



**UNIVERSITÀ
DEGLI STUDI
DI TRIESTE**



**UNIVERSITÀ
DEGLI STUDI
DI UDINE**
hic sunt futura

UNIVERSITÀ DEGLI STUDI DI TRIESTE UNIVERSITÀ DEGLI STUDI DI UDINE

XXXVI CICLO DEL DOTTORATO DI RICERCA IN INGEGNERIA CIVILE-AMBIENTALE E ARCHITETTURA

interateneo tra l'Università degli Studi di Trieste e l'Università degli Studi di Udine

APPROCCI MACHINE LEARNING PER LA MODELLAZIONE DEL COMPORTAMENTO FISICO- MECCANICO DI MISCELE BITUMINOSE PER INFRASTRUTTURE VIARIE

Settore scientifico-disciplinare: ICAR/04

**DOTTORANDO
FABIO RONDINELLA**

**COORDINATORE
PROF. ALBERTO SDEGNO**

**SUPERVISORE DI TESI
PROF. NICOLA BALDO**

**CO-SUPERVISORE DI TESI
PROF. CLARA CELAURO**

Prof. Ing. Clara Celauro

ANNO ACCADEMICO 2022/2023

Ai miei genitori

Sommario

Negli ultimi anni, uno dei concetti fondamentali che rappresenta la direzione verso cui tende la ricerca in molti settori dell'ingegneria è rappresentato dalla doppia transizione. Per doppia transizione si intende la fusione degli avanzamenti tecnologici sia in chiave digitale che ambientale al fine di soddisfare gli obiettivi di sviluppo sostenibile annunciati dalle Nazioni Unite. In questo contesto, gioca un ruolo cruciale l'ingegneria delle pavimentazioni in quanto settore fondamentale per lo sviluppo socio-economico di un Paese. A tal proposito, questo elaborato di tesi si propone di implementare innovativi algoritmi di apprendimento automatico per l'elaborazione di ingenti quantità di dati e lo sviluppo di modelli predittivi del comportamento fisico-meccanico di miscele bituminose per pavimentazioni stradali e/o aeroportuali. Attraverso moderne procedure di pre-elaborazione, generazione sintetica di osservazioni, validazione incrociata ed ottimizzazione degli iperparametri è stato infatti possibile prevedere con elevata precisione e ragguardevole attendibilità uno o (simultaneamente) più parametri legati alle caratteristiche prestazionali delle miscele investigate, anche se molto diverse tra loro. In questo senso, l'impatto dell'apprendimento automatico sulla moderna ingegneria delle pavimentazioni può risultare tutt'altro che trascurabile, andando ad interessare diverse fasi procedurali che coinvolgono la progettazione, la costruzione e la manutenzione delle pavimentazioni stradali e/o aeroportuali. Nello specifico, un risultato tangibile del presente elaborato è rappresentato dalla possibilità di semplificare e velocizzare le convenzionali procedure di progettazione delle pavimentazioni, riducendo i carichi di lavoro in laboratorio e la successiva generazione di scarti. In questo modo, oltre ad una potenziale digitalizzazione dei dati di caratterizzazione prestazionale dei conglomerati bituminosi, si provvede ad aumentare la sostenibilità dei correlati processi produttivi, soddisfacendo entrambe le richieste in termini di doppia transizione.

Abstract

In recent years, one of the key concepts representing the direction toward which research is heading in many areas of engineering is represented by the twin transition. The twin transition refers to the merging of both digital and environmental technological breakthroughs in order to meet the sustainable development goals announced by the United Nations. In this framework, pavement engineering plays a crucial role as a key sector for the socio-economic development of a country. In this respect, this dissertation aims to implement innovative machine learning algorithms for the processing of large amounts of data and the subsequent development of predictive models of asphalt mixtures' physical-mechanical behavior for road and/or airport pavements. By means of modern preprocessing procedures, data augmentation techniques, cross-validation and hyperparameter optimization, it has been possible to predict with high accuracy and remarkable reliability one or (simultaneously) several parameters related to the physical-mechanical behavior of the investigated mixtures, even if they greatly differ from each other. In this sense, the impact of machine learning on modern pavement engineering may be far from negligible, affecting several procedural stages of road and/or airport pavement design, construction and maintenance. In particular, a specific result of the current dissertation is the feasibility of simplifying and speeding up conventional pavement design procedures, reducing laboratory workloads and the subsequent waste generation. In this way, in addition to a potential digitalization of data in terms of performance characterization of asphalt mixtures, the sustainability of the associated production processes would be increased, satisfying both requirements of a successful twin transition.

Indice

Sommario	i
Abstract	iii
1 Introduzione	1
2 Cenni di Tecnologia Stradale e Aeroportuale	4
2.1 Aggregati lapidei	6
2.1.1 Dimensioni e granulometria	6
2.1.2 Forma ed appiattimento	6
2.1.3 Equivalente in sabbia	7
2.1.4 Blu di metilene	8
2.1.5 Prova Los Angeles	8
2.1.6 Resistenza alla levigazione	9
2.2 Leganti bituminosi	9
2.2.1 Caratterizzazione empirica	10
2.2.2 Prova di penetrazione	10
2.2.3 Prova di rammollimento	10
2.2.4 Punto di infiammabilità	11
2.2.5 Solubilità in solventi organici	11
2.2.6 Punto di rottura Fraass	12
2.2.7 Prova di duttilità	12
2.2.8 Prova di recupero elastico	12
2.3 Conglomerati bituminosi	13
2.3.1 Requisiti generali	14
2.3.2 Requisiti empirici	15
2.3.3 Requisiti fondamentali	17
2.4 Valutazione strutturale delle pavimentazioni	20
3 Cenni di Intelligenza Artificiale e Machine Learning	22
3.1 Cos'è l'Intelligenza Artificiale?	22
3.1.1 Pensare in modo umano: l'approccio della modellazione cognitiva	23
3.1.2 Pensare in modo razionale: l'approccio delle "leggi del pensiero"	23

INDICE

3.1.3	Agire in modo razionale: l'approccio dell'agente razionale	24
3.1.4	Agire in modo umano: l'approccio del Test di Turing	24
3.2	Machine Learning	26
3.2.1	Apprendimento non supervisionato	27
3.2.2	Apprendimento per rinforzo	28
3.2.3	Apprendimento supervisionato	28
3.3	Data Analysis	30
3.3.1	Raccolta e descrizione dei dati	30
3.3.2	Analisi esplorativa	31
3.3.3	Analisi qualitativa	31
3.3.4	Preparazione ed integrazione	31
3.3.5	Gestione e normalizzazione	32
3.3.6	Estrazione e selezione delle caratteristiche	32
3.3.7	Modellazione	33
3.4	Compromesso Bias-Varianza	34
3.4.1	Cosa si intende per Bias	34
3.4.2	Cosa si intende per Varianza	34
3.4.3	Spiegazione analitica	34
3.4.4	Fenomeni di Underfitting ed Overfitting	35
3.5	Early Stopping	38
3.5.1	Criteri d'arresto	39
3.6	Convalida Incrociata	41
3.6.1	Leave-One-Out Cross-Validation	41
3.6.2	k-fold Cross-Validation	42
3.6.3	Compromesso Bias-Varianza nella convalida incrociata	43
3.7	Regolazione degli iperparametri	44
3.7.1	Ricerca a griglia	45
3.7.2	Ricerca casuale	46
3.7.3	Ricerca intelligente	46
3.7.4	Ottimizzazione Bayesiana	47
3.8	Valutazione della bontà di adattamento	48
3.8.1	Metriche di classificazione	49
3.8.2	Metriche di regressione	50
4	Algoritmi di Machine Learning	54
4.1	Regressione Lineare	54
4.1.1	Regressione Ridge	55
4.1.2	Regressione Lasso	55
4.2	Macchine a Vettori di Supporto	55
4.2.1	Kernel Trick	56
4.3	Reti Neurali Artificiali	58
4.3.1	Analogia biologica	58
4.3.2	Neurone artificiale	60

INDICE

4.3.3	Gestione delle non linearità	62
4.3.4	Architettura di una rete neurale	64
4.3.5	Principali architetture neurali	65
4.3.6	Percettrone multistrato	67
4.3.7	Algoritmo di retropropagazione	68
4.3.8	Regolazione dei pesi sinaptici dello strato di output	70
4.3.9	Regolazione dei pesi sinaptici degli strati intermedi	71
4.3.10	Algoritmo Levenberg-Marquardt	73
4.4	Alberi Decisionali	76
4.4.1	Contesto algoritmico	77
4.4.2	Criteri di divisione	79
4.4.3	Criteri di arresto	80
4.4.4	Metodi di potatura	80
4.4.5	Algoritmo CART	81
4.4.6	Modelli Ensemble	82
4.4.7	Gradient Boosting	83
4.4.8	CatBoost	84
5	Approccio di Ricerca	88
6	Capacità Portante di una Pista di Volo	90
6.1	Introduzione	90
6.2	Campagna Sperimentale	93
6.2.1	Indagini in situ	93
6.2.2	Parametri del bacino di deflessione - DBPs	95
6.2.3	Procedura di back-calculation	96
6.3	Modellazione predittiva con approcci neurali superficiali	96
6.3.1	Regolarizzazione Bayesiana	99
6.3.2	Data Augmentation	101
6.4	Discussione dei risultati	102
6.5	Conclusioni	106
7	Caratterizzazione Prestazionale di Conglomerati Bituminosi	108
7.1	Introduzione	108
7.2	Campagna Sperimentale	111
7.2.1	Materiali	111
7.2.2	Indagini di laboratorio	112
7.3	Modellazione predittiva con approcci neurali profondi	116
7.3.1	Regolarizzazione L2	116
7.3.2	Ottimizzazione Bayesiana degli iperparametri	116
7.4	Discussione dei risultati	118
7.5	Conclusioni	122

8	Influenza di Filler Sostenibili sul Comportamento Meccanico di Miscele Bituminose	124
8.1	Introduzione	124
8.2	Campagna Sperimentale	126
8.2.1	Materiali	126
8.2.2	Indagini di laboratorio	127
8.3	Modellazione predittiva con approccio CatBoost	132
8.3.1	Data Augmentation	133
8.4	Discussione dei risultati	134
8.4.1	Risultati di laboratorio	134
8.4.2	Risultati della modellazione predittiva	139
8.5	Conclusioni	143
9	Modulo Complesso di Conglomerati Bituminosi	147
9.1	Introduzione	147
9.2	Campagna Sperimentale	149
9.2.1	Selezione dei materiali e test di modulo dinamico	149
9.2.2	Equazioni empiriche per $ E^* $	150
9.3	Modellazione predittiva con approccio CatBoost	153
9.3.1	Variabili di input ed output	155
9.3.2	Correlazioni esistenti tra le variabili	155
9.4	Discussione dei risultati	156
9.4.1	Analisi di sensitività	160
9.4.2	Analisi SHAP	161
9.4.3	Confronto tra modelli	162
9.5	Conclusioni	164
10	Conclusioni	166
	Elenco delle figure	168
	Elenco delle tabelle	171
	Bibliografia	172
A	Lista delle Pubblicazioni	196
B	Elenco delle Abbreviazioni	200

Capitolo 1

Introduzione

” *There were five exabytes of information created between the dawn of civilization through 2003, but that much information is now created every two days.*

— Eric Schmidt

Alla luce del rapido sviluppo della società civile e del costante progresso del settore dell'ingegneria delle infrastrutture di trasporto, la comprensione e le esigenze funzionali delle pavimentazioni stradali e/o aeroportuali aumentano di giorno in giorno. Inoltre, la comparsa di "nuovi" materiali funzionali e lo sviluppo di discipline trasversali hanno fornito un importante sostegno alla progettazione ed alla costruzione di molteplici tipologie di pavimentazioni funzionali eco-compatibili, ampliando ulteriormente il campo di ricerca sui materiali. Pertanto, la sostenibilità delle procedure di progettazione unitamente alla costruzione di pavimentazioni cosiddette *environmental-friendly* rappresenta l'attuale direzione prioritaria cui tendono le moderne tecnologie stradali e aeroportuali [Zhang, 2019]. Per incentivare una simile transizione, nel 2015 le Nazioni Unite hanno annunciato i 17 obiettivi di sviluppo sostenibile (*Sustainable Development Goals* - SDGs) al fine di focalizzare l'attenzione di tutti gli Stati membri su molteplici questioni economiche, ambientali e di sostenibilità [Lee et al., 2016, Subramaniam et al., 2023]. In particolare, il nono SDG fa riferimento alla costruzione di infrastrutture resilienti, all'industrializzazione sostenibile ed alla riduzione del consumo di risorse non rinnovabili [Senadjki et al., 2022]. Il raggiungimento di questi obiettivi rappresenta un fattore molto interessante sia dal punto di vista della fattibilità tecnica sia, al tempo stesso, in un'ottica di transizione verso l'economia circolare pianificata per il 2030 [Huang et al., 2022, Khajuria et al., 2022].

Il settore dell'ingegneria delle pavimentazioni è sempre stato particolarmente sensibile alle questioni legate al concetto di sostenibilità, e per molti anni i ricercatori di tutto il mondo hanno dedicato i loro sforzi alla comprensione delle conseguenze scaturite dal riutilizzo dei cosiddetti materiali secondari [Behera and Mishra, 2012, Abreu et al., 2015, Cai and Liu, 2020, Wu and Montalvo, 2021, Li et al., 2021, Zhao and Yang, 2023]. Diverse soluzioni tecnologiche all'avanguardia sono state sviluppate per consentire il corretto riciclo di materiali giunti al termine della loro vita utile e di sottoprodotti industriali nella produzione delle miscele bituminose. Inoltre, con l'obiettivo di migliorare ulteriormente la sostenibilità di tali miscele e delle

procedure seguite ai fini della loro produzione, negli ultimi anni sono state investigate anche metodologie innovative relative alla modellazione predittiva del loro comportamento meccanico [Di Benedetto et al., 2004, Ceylan et al., 2008, Sakhaeifar et al., 2015, Cao et al., 2016]. Essere in grado di conoscere in anticipo le proprietà volumetriche oppure il comportamento fisico-meccanico dei conglomerati bituminosi sulla base di poche variabili compositive può infatti apportare innumerevoli vantaggi. A titolo esemplificativo: (a) la stima dei parametri che devono essere utilizzati all'interno delle convenzionali procedure di progettazione delle pavimentazioni stradali può risultare considerevolmente più rapida e precisa; (b) è possibile risparmiare ingenti quantità di tempo e denaro che sarebbero necessari per la preparazione e la produzione di campioni aggiuntivi al fine di effettuare i test prestazionali richiesti, limitando il consumo di materie prime aggiuntive e la conseguente produzione di rifiuti [Nemati and Dave, 2018]. Per questi motivi, molti ricercatori hanno iniziato ad esplorare diverse potenziali strategie predittive. In prima battuta, sono state sviluppate delle equazioni empiriche *ad hoc* ed alcune complesse equazioni costitutive [Bari, 2005, Giunta and Angela Pisano, 2006, Witczak et al., 2007, Pasetto and Baldo, 2015]. Più recentemente, con l'avvento dell'intelligenza artificiale e più precisamente dell'apprendimento automatico, sono stati sviluppati approcci *machine learning* capaci di elaborare ingenti quantità di dati ed accelerare così la transizione digitale in ciascun settore della conoscenza, compreso ovviamente quello dell'ingegneria delle pavimentazioni. Tali approcci sono stati in grado di identificare affidabili relazioni tra le variabili contemplate oltre che fornire previsioni sempre più accurate del comportamento fisico-meccanico dei conglomerati bituminosi in esame [Specht et al., 2007, Mirzahosseini et al., 2011, Androjić and Marović, 2017, Alrashydah and Abo-Qudais, 2018, Ziari et al., 2018, Majidifard et al., 2019, Ghafari et al., 2022, Barugahare et al., 2020, Baldo et al., 2022b, Phung et al., 2023].

Gli obiettivi principali del presente elaborato possono essere inquadrati all'interno del contesto appena delineato poiché nascono dalla volontà di sviluppare ed implementare innovativi modelli predittivi del comportamento fisico-meccanico di miscele bituminose utilizzate per applicazioni stradali e/o aeroportuali utilizzando innovativi approcci *machine learning*. In questo modo, i parametri fisico-meccanici potranno essere accuratamente predetti e le metodologie sviluppate potranno integrare le convenzionali procedure di progettazione delle pavimentazioni. Pertanto, ne deriverebbero procedimenti dall'aumentata sostenibilità, semplificati e velocizzati, che ridurrebbero i carichi di lavoro richiesti in laboratorio e la conseguente produzione di scarti.

L'elaborato è stato quindi strutturato nella maniera seguente: il Capitolo 2 fornisce al lettore alcuni cenni relativi alla tecnologia stradale e/o aeroportuale, descrivendo alcune tra le principali prove di laboratorio condotte sui conglomerati bituminosi e sui loro componenti oltre che alcune delle procedure di valutazione strutturale eseguite in situ sulle pavimentazioni esistenti. Il Capitolo 3, invece, si propone di descrivere sinteticamente i concetti di intelligenza artificiale e *machine learning*, facendo chiarezza sulle loro principali distinzioni e focalizzando l'attenzione sui concetti di apprendimento, di analisi dati e di regressione. Il Capitolo 4 approfondisce le principali metodologie di apprendimento automatico solitamente impiegate per affrontare problemi di regressione, descrivendo la struttura ed il funzionamento algoritmico di ciascuna tecnica. Maggiore spazio è stato destinato alle reti neurali artificiali ed agli alberi decisionali in virtù, oltre che della loro maggiore complessità, delle loro considerevoli potenzialità di utilizzo e delle entusiasmanti capacità predittive. Il Capitolo 5, attraverso una sintetica

descrizione cosiddetta *step-by-step* dell'approccio di ricerca seguito, intende fornire al lettore uno strumento di supporto per la comprensione di tutte le fasi che hanno coinvolto l'analisi e la modellazione predittiva mediante algoritmi di apprendimento automatico supervisionato dei fenomeni descritti nei capitoli successivi. Pertanto, i Capitoli 6, 7, 8 e 9 illustrano lo sviluppo e la successiva implementazione di molteplici modelli predittivi in grado di stimare i valori di uno o (simultaneamente) più parametri prestazionali legati al comportamento meccanico ed alla caratterizzazione volumetrica di conglomerati bituminosi utilizzati in applicazioni anche molto differenti tra loro. L'attenzione è stata focalizzata sui parametri maggiormente significativi nell'ambito della tecnologia delle pavimentazioni stradali e/o aeroportuali, spaziando dalle caratteristiche volumetriche (vuoti d'aria) alle principali variabili prestazionali, come ad esempio stabilità e scorrimento Marshall, modulo di rigidità, resistenza alla trazione indiretta, modulo dinamico ed angolo di fase. Infine, nel Capitolo 10, sono state sintetizzate le principali conclusioni che si possono trarre dalla ricerca descritta in questo elaborato, evidenziandone i limiti e gli aspetti che necessiteranno di ulteriori approfondimenti nel prossimo futuro.

Capitolo 2

Cenni di Tecnologia Stradale e Aeroportuale

” *Some of the best theorizing comes after collecting data because then you become aware of another reality.*

— **Robert J. Shiller**

In linea generale, una pavimentazione stradale e/o aeroportuale rappresenta una struttura che separa gli pneumatici del traffico veicolare dal terreno di sottofondo. Denominata anche sovrastruttura, una pavimentazione è solitamente costituita da una successione di strati dotati di differenti caratteristiche fisico-meccaniche, progettati in funzione del ruolo che dovranno assolvere e delle tipologie di sollecitazioni cui saranno sottoposti. Veicoli in transito ed agenti atmosferici, infatti, inducono delle sollecitazioni la cui intensità varia con la profondità in funzione della rigidità e dello spessore di ogni singolo strato. Pertanto, i principali compiti svolti da una pavimentazione correttamente progettata e realizzata saranno:

- l'ideale ripartizione delle tensioni trasmesse al sottofondo compatibilmente con la sua capacità portante, di modo che non si verifichino deformazioni permanenti del piano viabile;
- la garanzia di una superficie di rotolamento regolare, stabile e poco deformabile nel tempo che assicuri un'adeguata aderenza pneumatico-pavimentazione e, conseguentemente, un adeguato comfort di marcia;
- il corretto allontanamento delle acque meteoriche al fine di proteggere sovrastruttura e sottofondo dalle azioni di gelo e disgelo.

A seconda dei materiali impiegati per la realizzazione dei singoli strati e della loro disposizione all'interno della sovrastruttura, è possibile distinguere molteplici tipologie di pavimentazione: flessibili, semi-rigide, rigide, composite, modulari o in terra [Christopher et al., 2010]. Tuttavia, per motivi legati alla semplificazione della trattazione, in questo elaborato si affronterà esclusivamente il tema delle pavimentazioni flessibili, limitatamente agli obiettivi posti dalla presente

ricerca. Solitamente, procedendo dall'alto verso il basso all'interno di una pavimentazione flessibile, è possibile identificare (Figura 2.1):

- uno strato di usura in conglomerato bituminoso di elevata qualità, il cui compito principale è quello di garantire le idonee proprietà di regolarità ed aderenza, resistendo alle azioni tangenziali trasmesse dai veicoli e garantendo durabilità rispetto agli effetti indotti dagli agenti atmosferici;
- uno strato di collegamento (o *binder*), anch'esso realizzato in conglomerato bituminoso, che risente in maniera minore dei fenomeni di degrado e delle azioni tangenziali indotte dal traffico veicolare e che, pertanto, può essere realizzato utilizzando materiali dalla qualità leggermente inferiore;
- uno strato di base, sottoposto a compressione e flessione, che ha il compito di resistere alle sole sollecitazioni derivanti dai veicoli che si traducono, nella parte inferiore dello strato, in tensioni di trazione. Può essere, pertanto, realizzato in conglomerato bituminoso oppure misto bitumato/granulare;
- uno strato di fondazione, il cui compito principale è quello di ridistribuire le azioni verticali di compressione su un'area sufficientemente ampia (compatibilmente con la capacità portante del sottofondo), realizzato in misto granulare o con terre stabilizzate con legante idraulico o bituminoso.



Figura 2.1: Successione di strati in una pavimentazione flessibile.

Nel presente capitolo, verranno brevemente descritti i principali materiali utilizzati per la realizzazione dei conglomerati bituminosi, facendo cenno ad alcune delle prove di laboratorio eseguite ai fini della loro caratterizzazione fisico-meccanica. Tali prove hanno infatti permesso di registrare le osservazioni contenute nelle banche dati che hanno supportato l'attività di ricerca descritta in questo elaborato.

2.1 Aggregati lapidei

Gli aggregati lapidei sono particelle granulari di origine minerale che hanno origine dalle rocce (siano esse ignee, sedimentarie o metamorfiche) e possono trovarsi in depositi naturali oppure essere prodotti mediante estrazione da cave. In questo caso, si è soliti parlare di "aggregati primari", intendendo come "secondari" tutti quei prodotti di origine artificiale o da riciclo (scorie derivanti da processi industriali, detriti da demolizioni o rifiuti di varia natura), utilizzati in parziale o totale sostituzione degli aggregati lapidei convenzionali.

2.1.1 Dimensioni e granulometria

Da un punto di vista meramente dimensionale, gli aggregati sono solitamente distinti in grossi e fini. La letteratura tradizionale identifica come grossi tutti quegli aggregati trattenuti al vaglio di apertura pari a 4 mm e definisce come fini quelli passanti a tale vaglio ma trattenuti a quello di apertura pari a 0.063 mm. La frazione minerale passante anche a quest'ultimo vaglio (talvolta tale soglia è identificata come 0.075 mm) costituisce una componente molto particolare che prende il nome di *filler* (o riempitivo). La principale caratteristica che identifica una miscela di aggregati è rappresentata dalla distribuzione percentuale relativa in peso dei grani secondo le loro dimensioni, comunemente denominata distribuzione granulometrica. Tale proprietà ha un'importante influenza diretta sulle caratteristiche di lavorabilità, sulle proprietà volumetriche e sulla risposta meccanica del prodotto finale. I risultati di un'analisi granulometrica effettuata su una miscela di aggregati sono espressi in termini di un diagramma semi-logaritmico noto come curva granulometrica.

2.1.2 Forma ed appiattimento

Un'ulteriore caratteristica fondamentale inerente agli aggregati riguarda la loro forma: è preferibile, infatti, che questa sia cubica o sferoidale e che permetta di generare una struttura interna regolare del prodotto finale, con superfici di contatto ben distribuite. Per tale motivo, sono da evitare elementi caratterizzati da forme piatte e/o allungate che, inevitabilmente, determinano inadeguate caratteristiche volumetriche e meccaniche. Identificando come S la dimensione minima di un generico elemento lapideo e come L la sua dimensione massima, questo si classifica come regolare (o sferico) quando si verifica la condizione $L/S < 3$. Per quanto riguarda una miscela, è possibile ottenere un'indicazione del suo grado di regolarità mediante l'indice di forma (*Shape Index* - SI), definito come [UNI EN 933-4, 2008]:

$$SI = 100 \cdot \frac{M_2}{M_1} \quad (2.1.1)$$

avendo cura di estrarre un campione rappresentativo della miscela di aggregati ed avendo definito come M_2 la massa totale degli elementi irregolari e come M_1 la massa totale dell'intero campione.

Di particolare interesse è anche l'indice di appiattimento (*Flakiness Index* - FI), in quanto utilizzato come criterio di selezione degli aggregati grossi. Esso si basa su un principio analogo

a quello definito per l'indice di forma ed è definito come [UNI EN 933-3, 2012]:

$$FI = 100 \cdot \frac{M_2}{M_1} \quad (2.1.2)$$

avendo questa volta definito come M_2 la somma delle masse dei grani di ciascuna classe passante attraverso il corrispondente staccio a barre con apertura di larghezza pari a $D_i/2$ e come M_1 la massa totale dei grani appartenenti a ciascuna classe granulometrica d_i/D_i (Tabella 2.1).

Tabella 2.1: Classi dimensionali ed apertura degli stacci a barre.

Classe granulometrica d_i/D_i (mm)	Larghezza apertura dello staccio a barre (mm)
63/80	40 ± 0.5
50/63	31.5 ± 0.5
40/50	25 ± 0.4
31.5/40	20 ± 0.4
25/31.5	16 ± 0.4
20/25	12.5 ± 0.4
16/20	10 ± 0.2
12.5/16	8 ± 0.2
10/12.5	6.3 ± 0.2
8/10	5 ± 0.2
6.3/8	4 ± 0.15
5/6.3	3.15 ± 0.15
4/5	2.5 ± 0.15

2.1.3 Equivalente in sabbia

Questo test viene condotto per determinare con rapidità il grado di pulizia degli aggregati fini, ovvero la presenza più o meno elevata di materiale contaminante di natura organica e argillosa che riveste i granuli. Tale presenza, infatti, pregiudica la qualità dell'aggregato e lo caratterizza come "non pulito". Secondo la normativa [UNI EN 933-8, 2016], la prova dell'equivalente in sabbia viene condotta sugli aggregati passanti al vaglio da 2 mm. Da un campione rappresentativo, una massa di 120 g di materiale essiccato viene posta in un cilindro trasparente graduato. Si aggiunge al cilindro una quantità di soluzione di cloruro di calcio fino a raggiungere circa 100 mm. Il contenuto del cilindro viene lasciato indisturbato per circa 10 minuti e poi, dopo aver inserito un tappo, il cilindro viene agitato manualmente (più di 90 volte) o in un agitatore meccanico per 30 ± 1 s. Successivamente, si aggiunge altra soluzione di cloruro di calcio al cilindro fino a riempirlo in corrispondenza dell'altezza di $380 \pm 0,25$ mm. Il cilindro e il suo contenuto vengono quindi lasciati indisturbati per 20 minuti. Durante questo periodo, la frazione sabbiosa pulita si deposita sul fondo della provetta mentre la frazione argillosa che ha reagito con la soluzione flocculante rimane in sospensione, formando due strati ben distinti. Vengono dunque rilevate l'altezza della sospensione (h_1) e l'altezza del materiale sedimentato

(h_2). Pertanto, è possibile ottenere il valore dell'equivalente in sabbia (*Sand Equivalent* - SE) attraverso la seguente formula:

$$SE = 100 \cdot \frac{h_2}{h_1} \quad (2.1.3)$$

2.1.4 Blu di metilene

Il test del blu di metilene viene utilizzato per determinare la qualità degli aggregati fini determinando la presenza di minerali argillosi attivi. A differenza dei minerali argillosi inattivi, le argille attive tendono a gonfiarsi a seconda del loro contenuto d'acqua. Questo rigonfiamento ha un impatto negativo sia sulla miscela bituminosa sia sugli strati non legati della pavimentazione. Il test del blu di metilene è integrativo del test dell'equivalente in sabbia, poiché quest'ultimo determina solo la presenza di particelle di argilla ma non quella di minerali argillosi attivi. Secondo quanto prescritto dalla normativa [UNI EN 933-9, 2022], il test del blu di metilene viene condotto facendo reagire una sospensione in acqua distillata della frazione di aggregati da analizzare con una soluzione colorante di blu di metilene in concentrazione predefinita. Successivamente, si preleva una goccia di sospensione e la si deposita su carta filtro. La macchia formata si presenta di un colore blu intenso circondata da una zona umida incolore. Il test è considerato positivo se, nella zona umida, si forma un alone costituito da un anello di colore blu chiaro persistente di circa 1 mm. La persistenza di tale alone è correlata all'assorbimento della soluzione colorante da parte dei minerali argillosi attivi. Pertanto, ripetendo la procedura fino al raggiungimento della condizione di persistenza, il volume complessivo della soluzione iniettata (V) in millilitri è legato alla quantità di materiale inquinante presente nell'aggregato. Esprimendo la massa del campione (M) in grammi, il valore del blu di metilene (*Methylene Blue Value* - MBV) è determinato tramite la seguente equazione:

$$MBV = 10 \cdot \frac{V}{M} \quad (2.1.4)$$

2.1.5 Prova Los Angeles

Tradizionalmente, una dei test maggiormente indicati al fine di caratterizzare il comportamento meccanico degli aggregati sottoposti all'azione perturbatrice del traffico e degli effetti termici è la cosiddetta prova di resistenza alla frammentazione. Tale prova si basa sull'utilizzo dell'apparecchiatura Los Angeles e ha come obiettivo la valutazione del degrado prodotto sugli aggregati grossi da parte dell'azione combinata di abrasione ed urto. Questa si sviluppa durante la rotazione dell'aggregato e di alcune sfere d'acciaio poste all'interno dell'apparecchiatura Los Angeles. La macchina è costituita da un tamburo rotante con dimensioni interne di 508 mm di lunghezza per 711 mm di diametro. Le dimensioni standard dell'aggregato testato in conformità alla normativa [UNI EN 1097-2, 2020] sono di 10/14 mm. La massa del campione di prova utilizzato è di $5000 \pm 5g$ ed è ottenuta da una massa campionata di almeno 15 kg di aggregato con la richiesta che il 60-70% della massa del campione di prova passi attraverso il vaglio da 12,5 mm. All'aggregato secco vengono aggiunte 11 sfere d'acciaio di diametro e massa noti. Il tamburo viene quindi chiuso e fatto ruotare 500 volte a velocità prefissata. Al termine del test, gli aggregati vengono setacciati per determinare la massa di materiale trattenuta dal vaglio

da 1.6 mm. Indicando con M_0 la massa iniziale del campione e con $M_{1.6}$ la massa trattenuta al vaglio di riferimento, il coefficiente Los Angeles (LA) è calcolato come:

$$LA = 100 \cdot \frac{M_0 - M_{1.6}}{M_0} \quad (2.1.5)$$

Ne consegue che quanto maggiore è il valore di LA, tanto minore sarà la qualità dell'aggregato di prova dal punto di vista meccanico e viceversa.

2.1.6 Resistenza alla levigazione

Questo test viene condotto sugli aggregati utilizzati per realizzare gli strati superficiali delle pavimentazioni stradali e determina la resistenza dell'aggregato all'azione di levigazione esercitata dal traffico veicolare. Il test si compone di due momenti: in una prima fase, i provini sono sottoposti a un'azione di levigazione in una macchina per la levigazione accelerata; in una seconda fase, invece, lo stato di levigazione raggiunto da ciascun provino viene misurato utilizzando un'apparecchiatura nota come pendolo britannico. All'atto pratico si procede utilizzando una resina per incollare l'aggregato opportunamente selezionato mediante criteri granulometrici su alcune lamelle ricurve alloggiare poi su di una ruota porta-provini. Questa viene fatta ruotare (ad una certa velocità e per un tempo prestabilito) essendo posta a contatto con una ruota in gomma piena che applica una pressione prevista da normativa [UNI EN 1097-8, 2020] e ruota in senso contrario. All'interfaccia tra campioni e ruotino viene indirizzata una carica abrasiva standardizzata composta da una miscela di acqua e sabbia calibrata. Simultaneamente alla prova in oggetto, è opportuno sottoporre al medesimo trattamento un materiale assunto come riferimento e dotato di caratteristiche standard. Il valore del parametro di levigabilità (*Polished Stone Value* - PSV) si ottiene mediante l'equazione:

$$PSV = S + 52.5 - C \quad (2.1.6)$$

avendo cura di indicare con S la media dei valori ottenuti (successivamente al trattamento e solitamente impiegando quattro provini) in seguito all'utilizzo di un pendolo britannico, e con C la medesima grandezza ottenuta per l'aggregato impiegato come riferimento.

2.2 Leganti bituminosi

Per legante bituminoso si intende generalmente un composto di natura organica derivato dal petrolio che è capace di conferire coesione ad un ammasso sciolto. A seconda della composizione iniziale della materia organica e delle condizioni prevalenti, il greggio di petrolio acquisisce le sue caratteristiche proprietà fisiche e la sua composizione chimica. Di conseguenza, esso varia da un giacimento all'altro e si può presentare con colori che vanno dal nero al marrone e con viscosità molto differenti tra loro. I bitumi tradizionali sono solitamente prodotti mediante distillazione frazionata del greggio di petrolio. Tuttavia, quelli utilizzati per la costruzione e la manutenzione di pavimentazioni stradali e/o aeroportuali, devono possedere particolari proprietà caratteristiche che andranno meticolosamente determinate mediante apposite procedure sperimentali ampiamente riconosciute dalla comunità scientifica.

2.2.1 Caratterizzazione empirica

Lo scopo delle prove di laboratorio eseguite sui bitumi è quello di definirne le proprietà caratteristiche, in modo da verificarne l' idoneità e prevederne il comportamento durante la vita utile della pavimentazione. Il termine "proprietà caratteristiche" comprende tutte le proprietà, come ad esempio quelle tecnologiche, quelle meccaniche, quelle reologiche e quelle fisico-chimiche. Le proprietà tecnologiche rappresentano quelle definite mediante prove empiriche e non su base razionale, a differenza di quanto accade per le proprietà meccaniche e reologiche. Pertanto, la caratterizzazione empirica manifesta importanti limitazioni sia di tipo concettuale che di tipo tecnologico-strumentale, risultando progressivamente meno efficace in virtù di una maggiore diversificazione dei prodotti e dell' utilizzo dei cosiddetti materiali innovativi. Inoltre, la natura dei test empirici risulta esclusivamente prescrittiva e non si presta ad una vera e propria qualificazione prestazionale. Alcune tra le principali proprietà tecnologiche sono rappresentate dalla penetrazione, dal punto di rammollimento, dal punto di infiammabilità, dalla solubilità in solventi organici e dal punto di rottura di Fraas. Tuttavia, le attuali normative europee di classificazione dei bitumi fanno ancora riferimento alle principali prove di caratterizzazione tecnologica. Pertanto, per motivi legati alla completezza della trattazione ed in virtù del fatto che tali parametri interessano alcune delle banche dati utilizzate per gli scopi di questa ricerca, si è deciso di riportare una sintetica descrizione di alcuni tra i principali protocolli relativi alle prove di caratterizzazione eseguite sui bitumi.

2.2.2 Prova di penetrazione

La prova di penetrazione rappresenta uno dei test più diffusi per la classificazione dei bitumi per pavimentazioni. Questa procedura determina anche la consistenza del bitume e dei leganti bituminosi ad una temperatura di esercizio intermedia (25°C). Valori più elevati di penetrazione indicano una consistenza più morbida mentre valori più bassi corrispondono ad una consistenza più dura. La prova consiste nel determinare, in condizioni standard di temperatura, carico applicato e durata dell' applicazione (5 secondi), la profondità di penetrazione di un ago metallico di peso e dimensioni predefinite all' interno del campione di bitume contenuto in un recipiente cilindrico standardizzato [UNI EN 1426, 2015]. Successivamente, la profondità di penetrazione dell' ago viene misurata in 0,1 mm o dmm.

2.2.3 Prova di rammollimento

La consistenza del bitume e dei leganti bituminosi a temperatura elevata viene determinata empiricamente mediante il test del punto di rammollimento, noto anche come test palla e anello (*Ring and Ball* - R&B). Insieme al test di penetrazione, il test del punto di rammollimento viene utilizzato per classificare i bitumi. Infatti, essendo un materiale viscoelastico, il bitume diventa gradualmente più morbido e meno viscoso all' aumentare della temperatura. Il punto di rammollimento è definito come la temperatura alla quale il materiale, in condizioni di prova standardizzate, raggiunge una determinata consistenza. Il test viene utilizzato per determinare il punto di rammollimento del bitume e dei leganti bituminosi nell' intervallo compreso tra 28°C e 150°C. Due anelli orizzontali di ottone riempiti di legante bituminoso sono montati su un

telaio metallico e riscaldati a velocità controllata in un bagno liquido mentre ciascuno sostiene una sfera di acciaio. Il punto di rammollimento è definito come la media delle temperature alle quali i due dischi rammolliscono abbastanza da permettere a ciascuna sfera, avvolta nel legante bituminoso, di cadere per una distanza di 25 mm. Il supporto con gli anelli contenenti il legante bituminoso viene posto in un recipiente di vetro riempito con acqua distillata o deionizzata (a circa 5°C), prima del posizionamento delle sfere di acciaio e dell'inizio della prova. Dopo 15 minuti, le sfere d'acciaio vengono posizionate su ciascun anello e l'acqua viene riscaldata ad una velocità uniforme di 5°C/min, seguendo le istruzioni della normativa [UNI EN 1427, 2015]. All'aumentare della temperatura, il bitume si ammorbidisce e inizia a formare un menisco che aumenta costantemente all'aumentare della temperatura (la cosiddetta forma a goccia). Per ogni palla e anello, occorre dunque registrare la temperatura indicata dal termometro nell'istante in cui il legante bituminoso che circonda la sfera tocca la piastra inferiore.

2.2.4 Punto di infiammabilità

I punti di infiammabilità e di combustione dei prodotti petroliferi, come i bitumi e i leganti bituminosi sono solitamente determinati tramite dispositivo Cleveland a coppa aperta. Il punto di infiammabilità rappresenta la temperatura più bassa alla quale l'applicazione di una fiamma di prova provoca l'accensione del vapore della porzione di prova e la propagazione della fiamma sulla superficie del liquido nelle condizioni specificate della prova. Il punto di combustione, invece, rappresenta la temperatura più bassa alla quale l'applicazione di una fiamma di prova provoca l'accensione del vapore della porzione di prova e la sua combustione per un minimo di 5 secondi nelle condizioni specificate dalla prova.

I punti di infiammabilità e di combustione sono utili per motivi di sicurezza e per evitare incidenti in caso di surriscaldamento del bitume. Secondo la normativa [UNI EN ISO 2592, 2017], una quantità di bitume che ha riempito la coppa di prova (coppa Cleveland) viene inizialmente riscaldata ad una velocità elevata (14÷17°C/min). Quando la temperatura sale a circa 56°C al di sotto del punto di infiammabilità previsto, si diminuisce il calore in modo che la velocità di progressivo aumento della temperatura per gli ultimi (23 ± 5)°C prima del previsto punto di infiammabilità sia di 5÷6°C/min. In questa fase di riscaldamento, una fiamma di prova passa ciclicamente, ogni 2°C, sul centro della coppa. La temperatura alla quale si verifica l'infiammabilità dei vapori è definita punto di infiammabilità. Il punto di combustione, invece, viene successivamente determinato utilizzando lo stesso dispositivo e proseguendo con il riscaldamento.

2.2.5 Solubilità in solventi organici

Il test di solubilità viene eseguito per determinare il grado di solubilità dei leganti bituminosi in un solvente specifico. Solitamente, il solvente in questione è il toluene, spesso utilizzato per i test di riferimento. Si noti che i leganti bituminosi avranno solubilità differenti nel caso in cui si utilizzino solventi differenti. Secondo la normativa [UNI EN 12592, 2007], una massa di 2 g di bitume viene sciolta in 100 ml di solvente e poi la soluzione viene filtrata. La quantità di sostanze trattenute dal filtro viene lavata, asciugata, pesata ed espressa come percentuale della massa iniziale di bitume. Questa percentuale, sottratta al 100%, determina la solubilità

del bitume. Il grado di solubilità deve essere superiore a un determinato valore, solitamente identificato pari al 99%.

2.2.6 Punto di rottura Fraass

Il punto di rottura Fraass fornisce una misura della fragilità del bitume e dei leganti bituminosi alle basse temperature (sotto zero). Il punto di rottura Fraass, secondo la normativa [UNI EN 12593, 2015], è la temperatura alla quale il legante bituminoso di uno spessore specifico e uniforme inizia a screpolarsi sotto condizioni di carico ben definite. L'apparecchiatura Fraass è costituita da uno strumento di piegatura in grado di applicare delle ripetute sollecitazioni di flessione, flettendo la piastra di prova rivestita (quindi sollecitando a trazione il film di bitume), da un sistema di raffreddamento, da un termometro graduato e da un'unità di preparazione del campione per creare il film di bitume sulla piastra di prova in acciaio. Tale piastra possiede dimensioni prestabilite e la quantità di legante bituminoso posta su di essa è di 0,4 g. A partire da una temperatura iniziale di 10°C, la piastra viene sottoposta a ripetute inflessioni con progressiva diminuzione della temperatura di 1°C al minuto. La temperatura alla quale compare la prima fessura sul film di bitume viene indicata come punto di rottura Fraass.

2.2.7 Prova di duttilità

La prova di duttilità misura indirettamente le proprietà di resistenza a trazione dei leganti bituminosi, in particolare di quelli modificati con polimeri. Viene eseguita mediante duttilometro, ovvero una vasca entro cui circola acqua alla temperatura di prova. Durante i test di duttilità, il provino di bitume, avente una caratteristica forma a clessidra, viene sottoposto ad allungamento ad una specifica velocità di deformazione (50 mm/min) e ad una specifica temperatura di prova finché non si rompe. La lunghezza di allungamento alla rottura, misurata in mm, è definita come valore di duttilità del bitume. Il test viene solitamente eseguito su tre campioni e la duttilità viene determinata sulla base della media dei tre valori. La temperatura di prova è generalmente di 5°C. Tuttavia, nel caso di leganti bituminosi morbidi, la prova può essere eseguita a una temperatura inferiore (0°C), mentre per i bitumi duri modificati con polimeri, la prova dovrebbe essere eseguita a 10°C o addirittura a 15°C [UNI EN 13589, 2018].

2.2.8 Prova di recupero elastico

La prova di recupero elastico viene utilizzata per determinare il recupero elastico dei leganti bituminosi in un duttilometro a una temperatura prestabilita. Si applica principalmente ai bitumi modificati con elastomeri termoplastici. Tuttavia, può essere utilizzata anche in presenza di bitumi convenzionali o di altri leganti bituminosi che generano solo un piccolo recupero. La prova viene eseguita in conformità alla normativa [UNI EN 13398, 2017], utilizzando la stessa apparecchiatura, lo stesso stampo e la stessa procedura della prova di duttilità. La temperatura di prova è solitamente di 25°C. L'unica differenza è che il provino non viene allungato fino alla rottura, ma solo fino al raggiungimento di un allungamento prestabilito (200 mm). Il cordoncino di bitume così prodotto viene tagliato al centro per ottenere due metà. Trascorso un tempo di recupero prestabilito (30 minuti), si misura l'accorciamento dei mezzi cordoncini e lo si esprime

come percentuale della lunghezza di allungamento. Pertanto, il recupero elastico (R_E) può essere espresso dalla seguente relazione:

$$R_E = 100 \cdot \frac{d}{200} \quad (2.2.1)$$

in cui d rappresenta la distanza tra le due porzioni di bitume soggetto a ritiro (mm) mentre R_E (%) esprime la percentuale di recupero elastico del bitume.

2.3 Conglomerati bituminosi

I conglomerati bituminosi svolgono un ruolo fondamentale nella costruzione e nella manutenzione di molte delle pavimentazioni stradali e/o aeroportuali in tutto il mondo. Questi sono ottenuti tramite la miscelazione e l'attenta combinazione di aggregati lapidei, legante bituminoso, filler ed eventuali additivi in grado di conferirgli le desiderate proprietà fisico-meccaniche che variano in funzione della composizione (assortimento granulometrico e dosaggio del legante), delle caratteristiche volumetriche e delle proprietà meccaniche dei componenti.

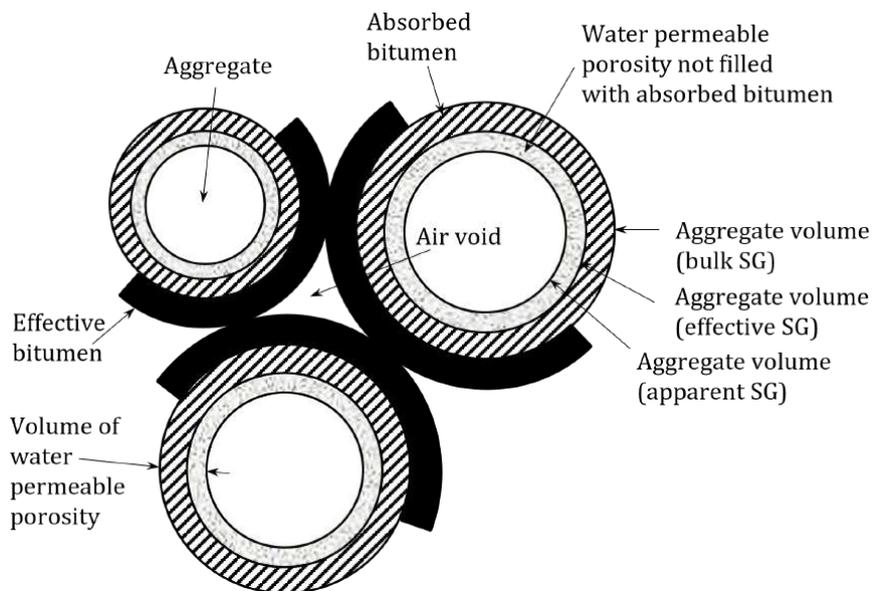


Figura 2.2: Composizione interna di un conglomerato bituminoso.

I conglomerati bituminosi prodotti a caldo sono progettati, all'interno dell'Unione Europea, in ottemperanza alla normativa [UNI EN 13108-1, 2016], la quale fornisce i requisiti generali dei materiali costituenti e del conglomerato stesso, nonché i criteri di valutazione della conformità. La normativa prevede due metodologie per definire le specifiche dei conglomerati bituminosi: il primo approccio, definito empirico, definisce i conglomerati mediante ricette compositive e requisiti per i materiali costituenti, accompagnati da requisiti aggiuntivi basati su prove prestazionali (approfonditi nei paragrafi 2.3.1 e 2.3.2); il secondo approccio, definito razionale,

definisce i conglomerati in termini di requisiti fondamentali legati ad una prescrizione limitata di composizione e materiali costituenti (approfonditi nei paragrafi 2.3.1 e 2.3.3).

2.3.1 Requisiti generali

Per quanto riguarda la miscela obiettivo, i requisiti generali prevedono la determinazione della composizione, della granulometria degli aggregati, del contenuto di vuoti, della resistenza alle deformazioni permanenti e della sensibilità all'acqua. Nel caso in cui il conglomerato debba essere impiegato in aree aeroportuali, è necessario determinare e dichiarare anche la sua resistenza al carburante ed ai liquidi antigelo. Per ragioni legate alla semplificazione della trattazione, soltanto alcuni dei presenti requisiti generali verranno discussi nelle sezioni successive.

Caratteristiche granulometriche degli aggregati

I requisiti della classificazione degli aggregati sono espressi in termini di percentuali massime e minime di passaggio attraverso alcuni vagli caratteristici: 1.4D, D, 2 mm e 0.063 mm. D rappresenta la dimensione nominale massima dell'aggregato espressa in mm e determina il grado del conglomerato. Una certa massa di aggregato può essere trattenuta da questo vaglio. 1.4D, invece, rappresenta il vaglio più vicino al risultato della moltiplicazione di D per un fattore 1.4, definendo l'apertura del vaglio attraverso cui deve passare tutta la massa dell'aggregato. I limiti granulometrici inferiore e superiore per ciascun vaglio caratteristico sono specificati in normativa [UNI EN 13108-1, 2016].

Contenuto di vuoti

Il contenuto di vuoti d'aria di ciascun provino compattato deve rientrare all'interno dei valori minimi e massimi definiti per le categorie indicate in normativa [UNI EN 13108-1, 2016]. Il contenuto minimo dei vuoti d'aria sarà compreso tra lo 0,5% ed il 6%, in funzione delle 12 diverse categorie. Allo stesso modo, il contenuto massimo dei vuoti d'aria dovrebbe essere compreso tra il 2% ed il 14%, in funzione delle 17 diverse categorie. La compattazione dei provini può essere effettuata mediante un compattatore a impatto (Marshall) secondo la normativa [UNI EN 12697-30, 2018], con un compattatore giratorio secondo la normativa [UNI EN 12697-31, 2019] o con un compattatore vibrante secondo la normativa [UNI EN 12697-32, 2019]. I provini possono essere realizzati anche mediante un compattatore a rulli, secondo la normativa [UNI EN 12697-33, 2019]. In tutti i casi, l'energia di compattazione applicata viene definita e dichiarata dal progettista ed il contenuto di vuoti d'aria (AV) viene calcolato attraverso la seguente equazione:

$$AV = 100 \cdot \frac{\rho_m - \rho_b}{\rho_m} \quad (2.3.1)$$

con ρ_m e ρ_b a rappresentare rispettivamente la massima densità del conglomerato bituminoso e la densità apparente del provino (g/cm^3).

Sensibilità all'acqua

La sensibilità all'acqua di una miscela bituminosa viene valutata determinando l'effetto della saturazione e del condizionamento accelerato in acqua sulla resistenza a trazione indiretta di provini cilindrici. Secondo la normativa [UNI EN 12697-12, 2018], occorre preparare una batteria di almeno sei provini cilindrici. Questi dovranno avere un diametro di 100 ± 3 mm qualora la dimensione nominale massima degli aggregati sia pari a 22 mm, oppure un diametro di 150 ± 3 mm o 160 ± 3 mm per dimensioni nominali massime maggiori. Successivamente, tali provini saranno soggetti a compattazione utilizzando uno dei metodi approvati e specificati nella normativa di riferimento. I provini saranno poi suddivisi in due sottogruppi di tre o più provini, aventi approssimativamente la stessa lunghezza media e la stessa densità media. Uno dei sottogruppi di provini viene conservato a $20^\circ\text{C} \pm 5^\circ\text{C}$ ed i provini di tale sottogruppo vengono denominati "asciutti". I provini dell'altro sottogruppo, invece, sono posti in un contenitore perforato all'interno di un recipiente in grado di realizzare il vuoto, riempito con acqua distillata a $20^\circ\text{C} \pm 5^\circ\text{C}$ fino ad un livello di almeno 20 mm superiore alla superficie superiore dei campioni. Si applica un vuoto di $6,7 \pm 0,3$ kPa per 30 ± 5 min, dopodiché i campioni restano immersi in acqua alla ristabilita pressione atmosferica per altri 30 ± 5 min. I provini di questo secondo sottogruppo vengono denominati "bagnati". Successivamente, il sottogruppo dei provini bagnati viene immerso in un bagno termostatico condizionato a $40^\circ\text{C} \pm 1^\circ\text{C}$ per un arco temporale compreso tra 68 e 72 ore. Successivamente al condizionamento, si determina la resistenza a trazione indiretta dei provini appartenenti ai due sottogruppi secondo la normativa [UNI EN 12697-23, 2018] definendo il rapporto di resistenza a trazione indiretta (*Indirect Tensile Strength Ratio* - ITS_R) come:

$$\text{ITS}_R = 100 \cdot \frac{\text{ITS}_w}{\text{ITS}_d} \quad (2.3.2)$$

avendo cura di indicare con ITS_w e con ITS_d rispettivamente la resistenza a trazione indiretta media dei provini bagnati e di quelli asciutti. Da normativa [UNI EN 12697-23, 2018], la resistenza a trazione indiretta (*Indirect Tensile Strength* - ITS) di ciascun provino si ottiene mediante l'equazione:

$$\text{ITS} = \frac{2 \cdot P}{\pi \cdot d \cdot t} \quad (2.3.3)$$

avendo definito come P il carico massimo e con le lettere d e t rispettivamente il diametro e l'altezza del provino. I requisiti generali in termini di sensibilità all'acqua prevedono che si raggiunga un valore minimo del rapporto di resistenza a trazione indiretta. Nello specifico, l'ITS_R minimo, espresso in percentuale, deve risultare pari al 60%, 70%, 80% oppure 90% in funzione delle 4 corrispondenti categorie.

2.3.2 Requisiti empirici

Oltre ai requisiti generali, la composizione obiettivo del conglomerato bituminoso deve soddisfare anche alcuni requisiti empirici, in termini di granulometria, contenuto minimo di legante, vuoti riempiti di bitume, vuoti nell'aggregato minerale e contenuto di vuoti a 10 giri di pressa giratoria. Se il conglomerato bituminoso è progettato per essere applicato in ambito aeroportuale, è necessario determinare anche i parametri Marshall. Analogamente a quanto fatto per i

requisiti generali, anche in questo caso alcuni dei presenti requisiti empirici saranno discussi nelle sezioni successive.

Vuoti nella miscela di aggregati

La minima percentuale dei vuoti nella miscela di aggregati (Voids in the Mineral Aggregate - VMA) dei provini compattati viene selezionata da opportune tabelle riportate in normativa [UNI EN 13108-1, 2016]. Il valori minimi di VMA variano tra l'8% ed il 16%, in funzione delle 6 corrispondenti categorie. I vuoti nella miscela di aggregati sono definiti attraverso la seguente equazione:

$$VMA = AV + B \cdot \frac{\rho_b}{\rho_B} \quad (2.3.4)$$

avendo definito B e ρ_B rispettivamente come il contenuto di legante della miscela (%) e la densità del legante stesso (g/cm^3).

Vuoti riempiti di bitume

Le percentuali minime e massime dei vuoti riempiti di bitume (Voids Filled with Bitumen - VFB) dei provini compattati vengono selezionate da opportune tabelle riportate in normativa [UNI EN 13108-1, 2016]. Quelle minime variano tra il 50% ed il 78%, in funzione delle 8 categorie corrispondenti mentre quelle massime variano tra il 50% ed il 97%, in funzione delle 16 categorie corrispondenti. I vuoti riempiti di bitume sono definiti attraverso la seguente equazione:

$$VFB = 100 \cdot \frac{B \cdot \frac{\rho_b}{\rho_B}}{VMA} \quad (2.3.5)$$

avendo mantenuto la coerenza dei simboli fino ad ora utilizzati.

Parametri Marshall

Sebbene la metodologia Marshall per la progettazione dei conglomerati bituminosi (*mix design*) sia piuttosto datata, essa è attualmente utilizzata da diverse aziende e/o istituzioni perché si avvale di attrezzature di laboratorio più semplici e meno costose. Tuttavia, pur non restituendo proprietà fondamentali del conglomerato in esame (come ad esempio la resistenza alle deformazioni permanenti), i parametri che si determinano mediante la prova Marshall forniscono un'idea della resistenza meccanica offerta dalla miscela di conglomerato sotto l'azione dei carichi dovuti al traffico [Santagata et al., 2009]. In generale, il *mix design* Marshall viene utilizzato per determinare il contenuto ottimale di bitume nei conglomerati bituminosi addensati contenenti aggregati di dimensione nominale massima pari a 25.4 mm. L'aggregato grosso, quello fine ed il filler vengono proporzionati in maniera tale da soddisfare i requisiti previsti da normativa. Successivamente, la quantità di miscela necessaria viene prelevata in modo da produrre provini di conglomerato bituminoso compattati del diametro di 101.6 mm. Per ottenere lo spessore desiderato (compreso tra 61.3 e 68.7 mm) sono necessari circa 1200 g di aggregati e filler. Gli aggregati vengono quindi riscaldati a una temperatura compresa tra 175° e 190°C mentre lo stampo di compattazione ed il martello vengono puliti e mantenuti preriscaldati a una temperatura compresa tra 100°C e i 145°C. Il bitume viene riscaldato ad una temperatura compresa tra

121°C e 138°C e la quantità necessaria di bitume di primo tentativo viene aggiunta all'aggregato riscaldato e mescolata accuratamente. La miscela viene posta in uno stampo e compattata con il numero di colpi specificato (75 per ciascun lato). Il campione viene estratto dallo stampo dopo pochi minuti utilizzando un estrattore. Per eseguire la prova di stabilità, il provino viene immerso in un bagno d'acqua alla temperatura di $60^{\circ} \pm 1^{\circ}\text{C}$ per un periodo di 30 minuti. Viene quindi collocato nell'apparecchiatura Marshall e caricato ad una velocità di deformazione costante di 50.8 mm/min fino al raggiungimento del carico di rottura. Il valore massimo (che causa la rottura del provino) espresso in kN definisce la stabilità Marshall mentre la deformazione totale subita dal provino al momento della rottura definisce lo scorrimento Marshall. Tuttavia, gli esperti ritengono che il modo in cui i campioni vengono compattati durante la procedura Marshall non simuli in maniera corretta la reale compattazione che avviene in situ e che la stabilità Marshall non stimi adeguatamente la resistenza a taglio del conglomerato.

Il *mix design* Marshall non è l'unica metodologia di progettazione delle miscele: esiste infatti l'approccio Superpave (uno dei principali risultati del Programma americano di Ricerca Strategica sulle Autostrade, *Strategic Highway Research Program - SHRP*) che si ritiene, prima o poi, sostituirà completamente il *mix design* Marshall [Nikolaides, 2015]. Una differenza fondamentale tra la metodologia Superpave e quella Marshall riguarda la modalità di compattazione dei provini: infatti, nella metodologia di *mix design* Superpave, i provini vengono compattati mediante compattatore giratorio che meglio simula gli effetti della compattazione reale in situ.

Nonostante queste premesse, i parametri Marshall come la stabilità, lo scorrimento ed il quoziente (conformemente alle normative [UNI EN 13108-20, 2016] e [UNI EN 12697-34, 2020]), costituiscono un requisito empirico quando il conglomerato bituminoso viene progettato per applicazioni aeroportuali. La normativa in materia di stabilità Marshall stabilisce requisiti non solo per un valore minimo, ma anche per un valore massimo [UNI EN 13108-1, 2016]. Nello specifico, i valori minimi di stabilità variano tra 2.5 e 12.5 kN, mentre quelli massimi variano tra 7.5 e 15.0 kN, in funzione delle 5 corrispondenti categorie. I valori di scorrimento variano tra 1 e 8 mm, in funzione delle 7 corrispondenti categorie. Infine, per quanto riguarda i valori del quoziente Marshall, i valori variano tra 1.0 e 4.0 kN/mm, in corrispondenza delle 7 categorie corrispondenti.

2.3.3 Requisiti fondamentali

Quando la composizione obiettivo di un conglomerato bituminoso è determinata sulla base dei requisiti fondamentali, questo deve soddisfare sia i requisiti generali, come indicato nella sezione 2.3.1, sia quelli fondamentali in termini di granulometria, di contenuto di legante, di rigidità, di resistenza alle deformazioni permanenti nella prova di compressione tri-assiale e di resistenza alla fatica.

Rigidità

La rigidità dei provini cilindrici deve essere compresa tra un valore minimo ed uno massimo, determinati in funzione dei valori e delle categorie previste dalla normativa [UNI EN 13108-1, 2016]. I valori di rigidità minima variano tra i 1500 e i 21000 MPa, in funzione delle 13 categorie corrispondenti. I valori di rigidità massima, invece, variano tra i 7000 e i 30000 MPa, in funzione

delle 8 categorie corrispondenti. I provini vengono preparati e compattati secondo la normativa [UNI EN 13108-20, 2016]. Successivamente, la rigidezza viene determinata utilizzando una delle procedure/apparecchiature descritte all'interno della normativa [UNI EN 12697-26, 2018] e secondo le condizioni di prova (temperatura e frequenza di carico) riportate nella tabella D.3 della normativa [UNI EN 13108-20, 2016]. Alcune delle menzionate procedure di prova verranno discusse nelle sezioni successive.

Trazione indiretta per provini cilindrici

Questa metodologia consente di misurare la rigidezza elastica delle miscele bituminose mediante prove di trazione indiretta, utilizzando provini cilindrici di vario diametro e spessore, prodotti in laboratorio o carotati da una pavimentazione stradale flessibile. La forma di carico è di tipo impulsivo e la prova viene eseguita in condizioni di deformazione controllata. Questa rappresenta probabilmente la prova più diffusa per la determinazione del modulo di rigidezza [Nikolaides, 2015]. Il diametro del provino da sottoporre al test può essere di 80, 100, 120, 150 o 200 mm e l'altezza può variare tra i 30 ed i 75 mm. Entrambe le dimensioni sono scelte in relazione alla dimensione nominale massima degli aggregati. Tuttavia, le dimensioni più comuni dei provini sono 100 o 150 mm di diametro e 40, 50 o 60 mm di altezza. Il provino cilindrico viene inserito all'interno di un telaio metallico ed il carico a impulsi viene applicato in modo da provocare una deformazione orizzontale di picco pari allo 0.005% del diametro del provino (per un provino di 100 mm di diametro, la deformazione orizzontale di picco risultante è di circa 5 μm , mentre per un provino di 150 mm di diametro il valore corrispondente è di circa 7 μm). Il tempo necessario affinché il carico applicato raggiunga il valore massimo, noto come tempo di crescita, è impostato a 124 ± 4 ms mentre i cicli di carico hanno un periodo pari a 3 ± 0.1 s. La prova viene generalmente eseguita a una temperatura di $20^\circ\text{C} \pm 0.5^\circ\text{C}$, ma è possibile utilizzare qualsiasi altra temperatura, solitamente inferiore a 20°C . La temperatura durante la prova deve essere mantenuta costante; pertanto, il dispositivo di prova viene collocato in un'apposita camera climatica. Dopo aver posizionato il provino e prima di iniziare a registrare i risultati, vengono applicati almeno 10 impulsi di carico di condizionamento (pre-carico), per consentire all'apparecchiatura di regolare l'entità e la durata del carico per ottenere la deformazione orizzontale specificata. Dopo gli impulsi di carico di condizionamento, vengono applicati altri cinque impulsi di carico, durante i quali vengono registrati il carico applicato e la deformazione orizzontale risultante. Per ogni impulso di carico, il modulo di rigidezza viene determinato con la seguente equazione:

$$E = \frac{F \cdot (\nu + 0.27)}{z \cdot h} \quad (2.3.6)$$

in cui E rappresenta il modulo di rigidezza (MPa) anche definito ITSM (*Indirect Tensile Stiffness Modulus*), F il valore di picco del carico verticale applicato (N), ν il coefficiente di Poisson, z l'ampiezza della deformazione orizzontale ottenuta durante il ciclo di carico (mm) e h lo spessore medio del provino (mm). Dopo aver completato la procedura descritta, il provino viene ruotato di $90^\circ \pm 10^\circ$ e si ripetono le stesse operazioni per la determinazione del modulo di rigidezza in questa nuova posizione. Il modulo di rigidezza del provino corrisponde al valore

medio delle due misurazioni, a condizione che la seconda misurazione sia compresa tra il +10% e -20% del valore medio registrato per la prima misurazione.

Flessione a quattro punti per provini prismatici

Con questa metodologia di prova, il modulo di rigidezza delle miscele bituminose viene misurato utilizzando un test a flessione. Il provino prismatico è sottoposto ad una flessione periodica a quattro punti con rotazione e traslazione (orizzontale) libere in tutti i punti di carico e reazione. La flessione è ottenuta spostando il punto di carico centrale in direzione verticale, perpendicolarmente all'asse longitudinale del provino. Le due estremità del travetto rimangono fisse. La forma di carico periodica è di tipo sinusoidale, al fine di ottenere l'ampiezza di deformazione richiesta di 50 ± 3 microstrain (ϵ). Secondo la normativa [UNI EN 12697-26, 2018], il provino ha la forma di un travetto prismatico con una lunghezza totale che non superi la lunghezza effettiva per più del 10%. La lunghezza effettiva (L) non deve essere inferiore a sei volte il valore massimo tra la larghezza (B) e l'altezza (H) del provino. La larghezza e l'altezza, inoltre, devono essere pari ad almeno tre volte la dimensione massima dei grani (D) presenti all'interno della miscela bituminosa in esame. Secondo la normativa [UNI EN 12697-26, 2018] (Allegato A) i provini sono prodotti mediante sezionamento da lastre realizzate in laboratorio secondo la norma [UNI EN 12697-33, 2019] o da lastre estratte da superfici stradali con spessore ≥ 60 mm. Durante la prova, la forza necessaria per la deformazione del provino viene misurata in funzione del tempo. La forza (F), la deflessione (z) e l'angolo di fase (φ) vengono registrati insieme alla temperatura e alla frequenza della prova. Il modulo di rigidezza iniziale è determinato come il modulo per un ciclo di carico compreso tra la 45-esima e la 100-esima ripetizione del carico (tipicamente determinato alla 100-esima ripetizione del carico). Secondo la normativa [UNI EN 12697-26, 2018], il modulo complesso E^* è definito come il rapporto tra sforzo e deformazione per un materiale viscoelastico lineare sottoposto a una forma d'onda di carico sinusoidale al tempo t , in cui lo sforzo applicato ($\sigma \cdot \sin(\omega t)$) risulta in una deformazione ($\epsilon \cdot \sin(\omega(t - \varphi))$) che ha un angolo di fase φ rispetto allo sforzo (con ω a rappresentare la velocità angolare espressa in rad/s). In questo caso, il modulo complesso (E^*) alla temperatura e alla frequenza richieste viene determinato mediante le equazioni:

$$|E^*| = \sqrt{(E_1^2 + E_2^2)} \quad (2.3.7)$$

$$\varphi = \arctan(E_2/E_1) \quad (2.3.8)$$

$$E_1 = \gamma \cdot [(F/z) \cdot \cos\varphi + 10^{-6} \cdot \mu \cdot \omega^2] \quad (2.3.9)$$

$$E_2 = \gamma \cdot (F/z) \cdot \sin\varphi \quad (2.3.10)$$

in cui γ rappresenta il fattore di forma (funzione delle dimensioni del campione e della sua forma) mentre μ rappresenta il fattore di massa (che tiene conto degli effetti inerziali dovuti alla massa del provino e delle parti mobili). I fattori di forma e di massa vengono determinati sulla base delle equazioni riportate in normativa [UNI EN 12697-26, 2018]. Le condizioni di prova tipicamente adottate in termini di temperatura e frequenza di carico corrispondono rispettivamente a 20°C e 8 Hz [Nikolaides, 2015].

2.4 Valutazione strutturale delle pavimentazioni

La valutazione strutturale delle pavimentazioni viene effettuata principalmente mediante dispositivi opportunamente progettati per eseguire prove non distruttive. Tuttavia, al fine di poter verificare la validità dei risultati ottenuti, risulta spesso necessario integrare tale valutazione eseguendo anche delle prove distruttive. Le prove distruttive generalmente comprendono l'estrazione di carote e la realizzazione di trincee esplorative. La prima procedura si rivela fondamentale per determinare (a) lo spessore degli strati di conglomerato bituminoso, (b) il grado di compattazione, (c) la composizione del conglomerato, (d) le proprietà meccaniche del conglomerato e (e) la presenza e l'eventuale estensione di fessurazioni all'interno degli strati del conglomerato in esame. La seconda, invece, si rende necessaria per (a) definire il comportamento meccanico di tutti gli strati al di sotto della pavimentazione, (b) determinare lo spessore ed il grado di compattazione di tali strati, (c) valutare il contenuto d'acqua, (d) definire la granulometria, (e) valutare la presenza di eventuali fessurazioni e (f) determinare le deformazioni degli strati. I dispositivi utilizzati per la valutazione strutturale non distruttiva delle pavimentazioni sono solitamente strumenti che impongono un carico alla superficie della pavimentazione e misurano la deflessione che ne consegue. La deflessione misurata, è la somma delle deflessioni di ciascuno strato, compreso il sottofondo, provocate dal carico di prova. Pertanto, l'entità della deflessione è un indicatore della capacità della pavimentazione di sopportare successivi carichi derivanti dal traffico. Infatti, maggiore è la deflessione che si verifica, minore è la capacità strutturale della pavimentazione di sopportare eventuali carichi successivi.

Dopo aver determinato la correlazione esistente tra la deflessione conseguente ad un tipo e ad un'intensità di carico ed il carico dell'asse standard equivalente (*Equivalent Single Axle Load* - ESAL), la valutazione strutturale della pavimentazione viene espressa in vita utile residua in termini di ESAL. Sulla base del valore dell'ESAL residuo, del traffico giornaliero medio espresso in ESAL e dell'incremento medio annuo del traffico, è possibile ottenere una stima della vita utile residua della pavimentazione in anni e/o determinare lo spessore di un eventuale ricoprimento necessario affinché la vita utile della pavimentazione aumenti di un certo numero di anni.

La valutazione strutturale non distruttiva di una pavimentazione può essere effettuata attraverso un'ampia gamma di dispositivi di misurazione della deflessione. Generalmente, tali dispositivi possono essere classificati in tre categorie:

- dispositivi statici, che devono arrestarsi per effettuare le misurazioni. Il carico può essere rotolante, vibrante o impulsivo. Ne sono un esempio la trave Benkelman o il deflettometro a massa battente;
- dispositivi semi-statici, che si muovono lentamente durante le misurazioni. Il carico è costituito da una massa rotolante o vibrante e l'apparecchiatura di misurazione è ferma e a contatto con la superficie della pavimentazione. Ne è un esempio il deflettografo LaCroix;
- dispositivi mobili, che effettuano misure di deflessione mentre sono in movimento. Il carico è rotolante e l'apparecchiatura di misurazione non è stazionaria e non è a contatto con la superficie della pavimentazione. Ne è un esempio il *Traffic Speed Deflectometer*.

Deflettometro a massa battente

Il deflettometro a massa battente (*Falling Weight Deflectometer* - FWD) è stato inizialmente sviluppato in Francia nel 1963 [Bretonnière, 1963] e poi perfezionato in Danimarca nel 1987 [Ullidtz, 1987] come dispositivo alternativo non distruttivo per la valutazione strutturale delle pavimentazioni. Oggi esistono molteplici FWD, tra i quali è possibile citare il Dynatest che successivamente verrà menzionato nel Capitolo 6. Tutti gli FWD si basano sullo stesso principio, ovvero applicano un determinato carico impulsivo utilizzando un sistema di caduta e sono tutti conformi alle linee guida dettate dalle normative [ASTM D4695-08, 2008, ASTM D4694-09, 2009]. Sono capaci di effettuare valutazioni strutturali di pavimentazioni stradali e/o aeroportuali, con i dispositivi impiegati in ambito aeroportuale in grado di esercitare carichi impulsivi dall'intensità maggiore e pertanto definiti *Heavy Weight Deflectometer* (HWD).

Un FWD è in grado di fornire informazioni sulle condizioni strutturali di tutti gli strati della pavimentazione, compreso il sottofondo. In particolare, con le misure di deflessione FWD è possibile ottenere: (a) la stima del modulo di elasticità o del modulo di rigidezza di ogni strato, (b) la stima della vita utile residua della pavimentazione, (c) l'individuazione dei punti più vulnerabili della pavimentazione, (d) la stima della severità delle potenziali fessure negli strati stabilizzati, (e) la determinazione dell'efficacia del trasferimento del carico ai giunti delle pavimentazioni rigide e (f) l'individuazione delle fessure nelle pavimentazioni rigide.

Nella maggior parte dei casi, il sistema FWD è montato su un rimorchio, piuttosto che su un veicolo. Il carico viene applicato attraverso una piastra metallica circolare di 300 mm di diametro per le pavimentazioni stradali oppure attraverso una piastra di 450 mm di diametro per le pavimentazioni aeroportuali. Il carico applicato per le convenzionali pavimentazioni flessibili stradali è solitamente di 50 kN. Le deflessioni risultanti vengono misurate al centro del carico applicato (attraverso un foro al centro della piastra di carico) e a varie distanze dal carico, mediante appositi sensori. I sensori possono essere trasduttori di spostamento, trasduttori di velocità, accelerometri, geofoni o anche sismometri. La deflessione viene misurata con l'ausilio di sette sensori posizionati su una linea retta passante per il centro dell'area di carico. Il primo sensore è posizionato esattamente al centro della piastra di carico (con un foro), mentre gli altri sono collocati a determinate distanze da esso. A titolo esemplificativo, la deflessione centrale fornisce un'indicazione delle prestazioni complessive della pavimentazione, mentre la deflessione a 1500 mm è un'indicazione delle condizioni del sottofondo.

Le misure vengono solitamente effettuate ogni 20 m o più, e ogni 200 m nel caso di rilievi su una rete stradale molto lunga. Quando le misurazioni vengono effettuate a grande distanza l'una dall'altra, in alcuni tratti può rendersi necessario effettuare misurazioni anche ogni 20 m, al fine di individuare con precisione la presunta criticità della pavimentazione. Poiché la temperatura influisce sul comportamento meccanico degli strati di conglomerato bituminoso, è opportuno considerare e misurare anche la temperatura della pavimentazione durante le prove. A tal fine, si rende necessario effettuare una misurazione a circa 100 mm sotto la superficie della pavimentazione, utilizzando un idoneo termometro elettronico. Successivamente, tutti i dati raccolti vengono elaborati mediante un software fornito dal produttore del FWD.

Capitolo 3

Cenni di Intelligenza Artificiale e Machine Learning

” *Artificial Intelligence is the study of how to make computers do things at which, at the moment, people are better.*

– Elain Rich

3.1 Cos'è l'Intelligenza Artificiale?

L'espressione Intelligenza Artificiale (*Artificial Intelligence - AI*) innegabilmente suscita emozioni. Da un lato troviamo il sostantivo *intelligenza* ed il fascino per questa qualità che conferisce agli esseri umani un posto speciale tra le forme di vita; dall'altro l'attributo *artificiale* che rimanda all'imitazione di fenomeni naturali tramite accorgimenti e/o procedimenti tecnici. In questo senso, risulta piuttosto complesso definire l'intelligenza artificiale in modo semplice ma rigoroso. Nel corso degli anni, si sono susseguite molteplici definizioni che possono essere riassunte come in Tabella 3.1 [Russell and Norvig, 2010].

Le definizioni contenute nella riga superiore riguardano i processi di pensiero e ragionamento, mentre quelle contenute nella riga inferiore fanno riferimento all'azione ed al comportamento. Inoltre, le definizioni contenute nella colonna di sinistra misurano il livello di successo in termini di fedeltà alle performance umane, mentre quelle contenute nella colonna di destra effettuano misure in termini di performance ideali, definite come razionalità. Un sistema si definisce razionale se fa la *cosa giusta* sulla base di ciò che sa. Storicamente, tutti e quattro gli approcci all'intelligenza artificiale sono stati seguiti, ognuno da diversi studiosi e con metodi differenti. Un approccio incentrato sull'uomo deve essere in parte una scienza empirica, che comporta osservazioni e ipotesi sul comportamento umano. L'approccio razionalista, invece, prevede una combinazione di matematica e ingegneria.

Tabella 3.1: Alcune definizioni di Intelligenza Artificiale organizzate per categorie.

Thinking Humanly	Thinking Rationally
“The exciting new effort to make computers think . . . <i>machines with minds</i> , in the full and literal sense.” [Haugeland, 1985] “[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . .” [Bellman, 1978]	“The study of mental faculties through the use of computational models.” [Charniak and McDermott, 1985] “The study of the computations that make it possible to perceive, reason, and act.” [Winston, 1992]
Acting Humanly	Acting Rationally
“The art of creating machines that perform functions that require intelligence when performed by people.” [Kurzweil, 1990] “The study of how to make computers do things at which, at the moment, people are better.” [Rich and Knight, 1991]	“Computational Intelligence is the study of the design of intelligent agents.” [Poole et al., 1998] “AI . . . is concerned with intelligent behavior in artifacts.” [Nilsson, 1998a]

3.1.1 Pensare in modo umano: l’approccio della modellazione cognitiva

Affinché un dato programma pensi come un essere umano, è necessario determinare in che modo pensino gli esseri umani e come funzionino effettivamente la loro mente. Esistono tre strumenti principali per raggiungere un simile obiettivo: l’introspezione, che cerca di cogliere i nostri pensieri mentre si susseguono; gli esperimenti psicologici, che osservano un essere umano in azione; e le moderne tecniche di visualizzazione (o *imaging*) cerebrale, che osservano il cervello in azione. Una volta ottenuta una teoria della mente sufficientemente precisa, è possibile esprimerla sotto forma di programma informatico. Il fatto che il comportamento di input-output del programma sia analogo al corrispondente comportamento umano, costituisce la prova che alcuni dei meccanismi operativi del programma in questione potrebbero somigliare a quelli presenti nella mente umana.

3.1.2 Pensare in modo razionale: l’approccio delle "leggi del pensiero"

Il filosofo greco Aristotele è stato uno dei primi a tentare di codificare il *giusto pensiero*, ovvero i processi corretti ed inconfutabili del ragionamento. I suoi sillogismi hanno fornito dei modelli di strutture argomentative in grado di produrre sempre conclusioni corrette qualora venissero fornite delle premesse corrette. Ad esempio: "Socrate è un uomo; tutti gli uomini sono mortali; quindi Socrate è mortale". Si suppone che queste leggi del pensiero governino il funzionamento della mente umana ed il loro studio ha dato origine al campo della logica. Nel XIX secolo i logici svilupparono una notazione precisa per le affermazioni circa tutti i tipi di oggetti del mondo e sulle relazioni tra di essi, in contrasto con la notazione aritmetica ordinaria che prevede solo affermazioni sui numeri. Già a partire dal 1965, esistevano programmi in grado di risolvere (in linea di principio) qualsiasi problema risolvibile descritto in notazione logica. La cosiddetta

tradizione logicista nell'ambito dell'intelligenza artificiale spera di basarsi su tali programmi per creare sistemi intelligenti. Ci sono due ostacoli principali a questo approccio. In primo luogo, non è semplice dichiarare la conoscenza informale nei termini formali richiesti dalla notazione logica, in particolare quando questa conoscenza non è sicura al 100%. In secondo luogo, c'è una grande differenza tra risolvere un problema "in linea teorica" e farlo in pratica. Anche problemi con poche centinaia di eventi possono esaurire le risorse computazionali di qualsiasi computer, a meno che questo non abbia una guida su quali processi di ragionamento tentare per primi. Sebbene entrambi questi ostacoli si applichino a qualsiasi tentativo di costruire sistemi di ragionamento computazionale, sono apparsi per primi nella tradizione logicista.

3.1.3 Agire in modo razionale: l'approccio dell'agente razionale

Un agente è semplicemente qualcosa che agisce (agente deriva dal latino *agere*, fare). Naturalmente, tutti i programmi per computer fanno qualcosa, ma ci si aspetta che gli agenti informatici facciano di più: operare in modo autonomo, percepire l'ambiente, persistere per un periodo di tempo prolungato, adattarsi al cambiamento, creare e perseguire obiettivi. Un agente razionale è un agente che agisce in modo da ottenere il miglior risultato o, in caso di incertezza, il miglior risultato previsto. Nell'approccio all'intelligenza artificiale basato sulle "leggi del pensiero", l'enfasi veniva posta sui ragionamenti corretti. Fare ragionamenti corretti fa parte dell'essere un agente razionale, perché agire razionalmente significa ragionare logicamente fino alla conclusione che una determinata azione permetterà di raggiungere i propri obiettivi, e poi agire in base a tale conclusione. D'altra parte, però, il ragionamento corretto non è tutto nella razionalità; in alcune situazioni, infatti, non c'è una cosa provatamente corretta da fare, sebbene qualcosa debba comunque essere fatto. L'apprendimento non serve solo per semplice erudizione, ma anche perché migliora la capacità di generare comportamenti efficaci. L'approccio dell'agente razionale presenta dunque due principali vantaggi rispetto agli altri approcci. In primo luogo, è più generale del pensiero razionale, poiché il ragionamento corretto è solo uno dei diversi meccanismi possibili per raggiungere la razionalità. In secondo luogo, è più adatto allo sviluppo scientifico rispetto agli approcci basati sul pensiero e sul comportamento umani. Lo standard di razionalità è matematicamente ben definito e completamente generale. Il comportamento umano, invece, è ben adattato a un ambiente specifico ed è definito dalla somma di tutte le cose che gli esseri umani fanno. Tuttavia, è importante tenere a mente che raggiungere la perfetta razionalità - fare sempre la cosa giusta - diventa impraticabile in ambienti complicati poiché i requisiti computazionali sarebbero troppo elevati.

3.1.4 Agire in modo umano: l'approccio del Test di Turing

Il Test di Turing [Turing, 1950] è stato concepito per fornire una soddisfacente definizione operativa dell'intelligenza. Un computer supera il test se un interrogatore umano, dopo aver posto alcune domande scritte, non riesce a distinguere se le risposte scritte provengano da una persona o da un computer. Il computer dovrebbe essere in grado di:

- elaborare un linguaggio naturale (*natural language processing*) che gli permetta di comunicare con successo in inglese;

- rappresentare la conoscenza (*knowledge representation*) per memorizzare ciò che ascolta o che sa;
- ragionare in modo automatizzato (*automated reasoning*) in modo da utilizzare le informazioni memorizzate per rispondere alle domande e per trarre conclusioni;
- apprendere automaticamente (*machine learning*) per adattarsi a nuove circostanze e per rilevare ed estrapolare modelli (Figura 3.1).

Il test di Turing evitava deliberatamente la diretta interazione fisica tra interrogatore e computer, poiché la simulazione fisica di una persona non è necessaria per definire l'intelligenza. Tuttavia, il cosiddetto Test di Turing Totale [Harnad, 1991] include un segnale video e la possibilità di far passare oggetti fisici attraverso una botola per testare sia le capacità percettive del soggetto che quelle di manipolazione degli oggetti. Per superare il Test di Turing Totale, il computer avrà bisogno di:

- visione computerizzata (*computer vision*) per percepire gli oggetti;
- automazione (*robotics*) per manipolare oggetti e muoversi.

Le sei discipline appena menzionate costituiscono gran parte dell'intelligenza artificiale e a Turing va il merito di aver ideato un test che rimane ancora attuale nonostante siano trascorsi più di settant'anni. Tuttavia, nel corso degli anni, i ricercatori nell'ambito dell'intelligenza artificiale hanno dedicato sempre meno attenzioni al superamento del test di Turing, ritenendo fosse più importante studiare i principi alla base dell'intelligenza piuttosto che duplicare un esemplare.

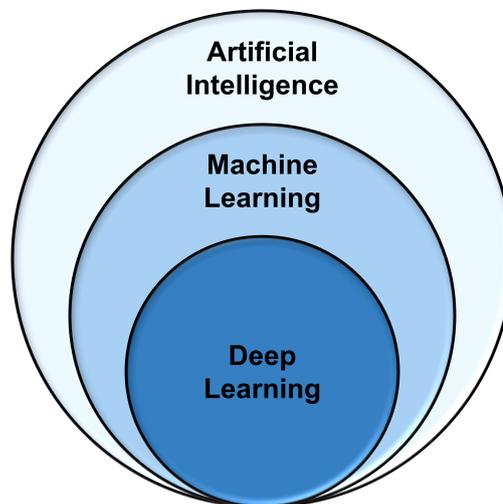


Figura 3.1: Rapporto tra AI, machine learning e deep learning.

3.2 Machine Learning

L'apprendimento, analogamente all'intelligenza, copre una così ampia gamma di processi che è difficile da definire con precisione. Facendo specifico riferimento alle macchine, si è soliti parlare di Apprendimento Automatico (*Machine Learning* - ML) ogni qual volta esse modificano la propria struttura o il proprio funzionamento sulla base di determinati input (o in risposta ad informazioni esterne) al fine di migliorare le proprie performance future [Nilsson, 1998b]. Ad esempio, quando le performance di una macchina per il riconoscimento vocale migliorano in seguito all'ascolto di diversi campioni audio della voce di una persona, diventa più che legittimo affermare che la macchina abbia imparato. Ci si potrebbe chiedere: "Perché le macchine devono imparare? Perché non progettarle in modo che funzionino sin dall'inizio come desiderato?". Innanzitutto, la messa a punto di macchine in grado di apprendere permette di comprendere più a fondo i meccanismi di apprendimento degli stessi esseri umani. Inoltre, dal punto di vista ingegneristico, esistono molteplici ulteriori motivazioni come quelle di seguito riportate.

- Alcune attività non possono essere definite correttamente se non attraverso un esempio. Infatti, è possibile che si sia in grado di specificare coppie di input/output ma non una precisa relazione tra gli input e gli output desiderati. Sarebbe dunque auspicabile che le macchine siano in grado di regolare la loro struttura interna in modo da produrre output corretti per un gran numero di input campione e quindi di vincolare opportunamente la loro funzione di input/output per approssimare la relazione implicita negli esempi;
- È possibile che nelle grandi banche dati si nascondano importanti relazioni e correlazioni: i metodi ML possono essere utilizzati per estrarre tali relazioni (*data mining*);
- I progettisti umani talvolta producono macchine che non funzionano esattamente come desiderato negli ambienti in cui vengono utilizzate. Ciò accade perché alcune caratteristiche dell'ambiente di lavoro potrebbero non essere completamente note al momento della progettazione. I metodi di apprendimento automatico possono essere dunque utilizzati per migliorare "sul posto" i progetti di macchine esistenti;
- Le conoscenze disponibili circa determinate attività potrebbero essere troppo vaste per essere codificate esplicitamente dall'uomo. Le macchine che apprendono tale conoscenza in modo graduale potrebbero essere in grado di catturare più di quanto l'uomo voglia scrivere;
- Gli ambienti mutano nel tempo. Le macchine in grado di adattarsi a un ambiente mutevole ridurrebbero la necessità di una costante riprogrammazione;
- Gli esseri umani scoprono continuamente nuove nozioni circa le attività che svolgono. C'è un flusso costante di nuovi eventi nel mondo. Anche il vocabolario cambia. La continua riprogrammazione dei sistemi AI per adeguarsi alle nuove conoscenze sarebbe impraticabile ma i metodi di apprendimento automatico potrebbero essere in grado di tenere traccia di gran parte di queste.

Qualsiasi componente di una macchina può essere migliorato imparando dai dati a disposizione. I miglioramenti e le tecniche utilizzate per ottenerli dipendono sostanzialmente da

quattro fattori principali: (1) componente da migliorare; (2) conoscenze pregresse possedute dalla macchina; (3) rappresentazione utilizzata per dati e componente; (4) *feedback* previsto per l'apprendimento. Focalizzando l'attenzione sul quarto fattore, esistono tre tipologie di *feedback* che determinano le tre principali strategie di apprendimento (Figura 3.2).

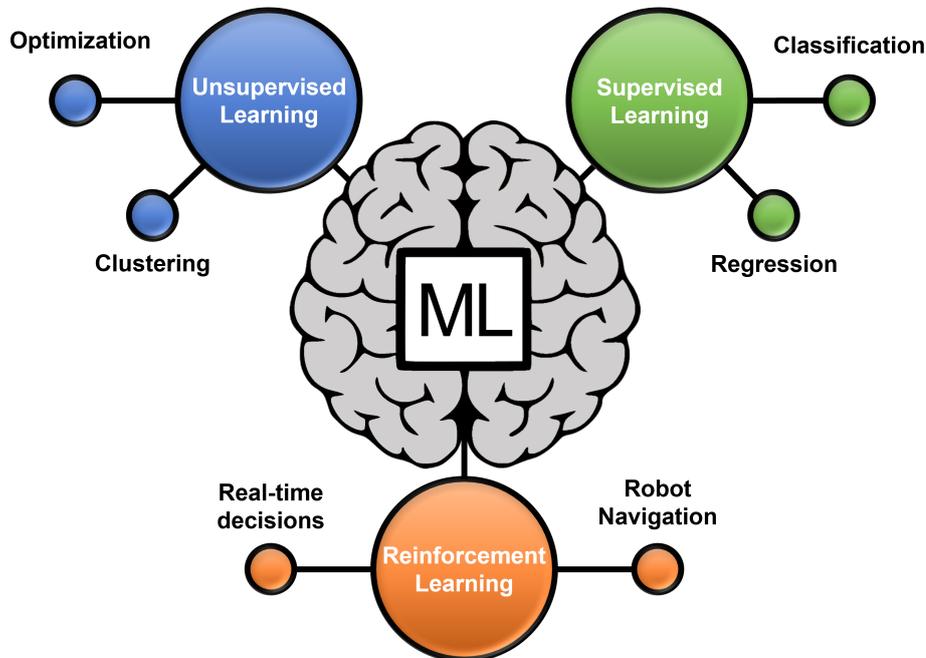


Figura 3.2: Principali strategie di apprendimento automatico ed esempi applicativi.

3.2.1 Apprendimento non supervisionato

Di seguito, viene riportata una breve discussione circa la prima tipologia di apprendimento automatico, ovvero quella di apprendimento non supervisionato. In questo caso, alla macchina non viene fornito alcun feedback esplicito cosicché essa possa apprendere schemi e modelli "liberamente". Il lavoro più comune svolto per mezzo di algoritmi di apprendimento non supervisionato prende il nome di raggruppamento (*clustering*) e consiste nel raggruppare oggetti in classi omogenee. Un *cluster* identifica un insieme di oggetti che presentano delle similarità tra loro e delle dissimilarità con oggetti presenti in altri *cluster*. In questo caso, i dati di partenza non sono categorizzati né etichettati in alcun modo. Altri compiti svolti tramite algoritmi di apprendimento non supervisionato prendono il nome di: rilevamento di anomalie (*anomaly detection*), riduzione della dimensionalità (*dimensionality reduction*) ed ottimizzazione (*optimization*).

3.2.2 Apprendimento per rinforzo

Tale strategia di apprendimento, così come nel caso di apprendimento non supervisionato, consiste nel fornire in input alla macchina dei dati non etichettati pertanto non collegati ad output corretti. Tuttavia, durante l'apprendimento per rinforzo, alla macchina viene fornito un *feedback*/rinforzo sulla base del quale poter imparare. Si tratta di premi o punizioni: i primi segnalano alla macchina che ha svolto correttamente un compito mentre le seconde le segnalano di aver commesso un errore. È compito della macchina decidere quali azioni precedenti al rinforzo siano state maggiormente determinanti. Tale strategia di apprendimento risulta particolarmente adatta a contesti in cui l'ambiente circostante è in grado di mutare poiché fornisce alla macchina gli strumenti per adattarsi a tali cambiamenti.

3.2.3 Apprendimento supervisionato

L'ultima strategia di apprendimento affrontata consiste nell'apprendimento supervisionato. In questo caso, la macchina ha la possibilità di osservare un set di dati etichettati di addestramento che, pertanto, comprende input ed output corretti. Tali coppie permettono alla stessa di apprendere una funzione di mappatura che, sulla base di nuovi input, le consente di produrre l'output desiderato. La macchina misura la precisione delle sue previsioni attraverso una funzione di perdita ed il processo di apprendimento continua permettendole di modificarsi fino a quando l'errore non sarà stato sufficientemente minimizzato. L'apprendimento supervisionato consente di affrontare due diverse tipologie di problemi:

- **Classificazione:** quando l'output è rappresentato da un insieme finito di valori (etichette - categorie). Prende il nome di classificazione Booleana (o binaria) se l'output è rappresentato da due soli possibili valori. A titolo esemplificativo, prevedere il vincitore di una gara sportiva (squadra A o squadra B) oppure determinare il genere di un determinato film (commedia, drammatico o d'animazione) si definiscono come problemi di classificazione poiché ognuno di questi comporta la selezione di un'etichetta tra le possibili alternative;
- **Regressione:** quando l'output è rappresentato da una quantità numerica. Prevedere il peso di una persona oppure quanta neve cadrà quest'anno rappresentano problemi di regressione, in cui si prevede il valore numerico futuro di una funzione a partire da valori precedenti ed altre caratteristiche rilevanti.

Una rappresentazione sintetica delle menzionate strategie di apprendimento viene riportata in Tabella 3.2 in cui sono descritti i principali vantaggi e svantaggi di ciascuna strategia.

Tabella 3.2: Vantaggi e svantaggi delle diverse strategie di apprendimento automatico.

Strategia di apprendimento automatico	Vantaggi	Svantaggi
Non supervisionato	È in grado di risolvere complessi problemi di ML ed è in grado di affrontare compiti multipli	Algoritmo complesso che, unitamente a dati non etichettati, può risultare in previsioni imprecise
Per rinforzo	È semplice sia da comprendere che da implementare e talvolta sopperisce agli inconvenienti delle altre strategie di apprendimento	I risultati delle iterazioni possono essere instabili e talvolta poco accurati
Supervisionato	Aiuta a prevedere i risultati in base all'esperienza pregressa ed a fornire un'idea precisa delle classi di oggetti	Non è ideale per risolvere problemi complessi e necessita di lunghi tempi computazionali per addestrare i modelli

3.3 Data Analysis

La risoluzione dei problemi di apprendimento automatico solitamente coinvolge un'ormai consolidata sequenza di passaggi associati alla raccolta dei dati, all'estrazione di informazioni, alla preparazione di un modello, alla valutazione delle performance ottenute ed alla sua implementazione [Harrington, 2012, Sarkar et al., 2018]. Tale sequenza può essere schematizzata come in Figura 3.3 [Subasi, 2020] e viene descritta nel dettaglio nei seguenti sottoparagrafi.

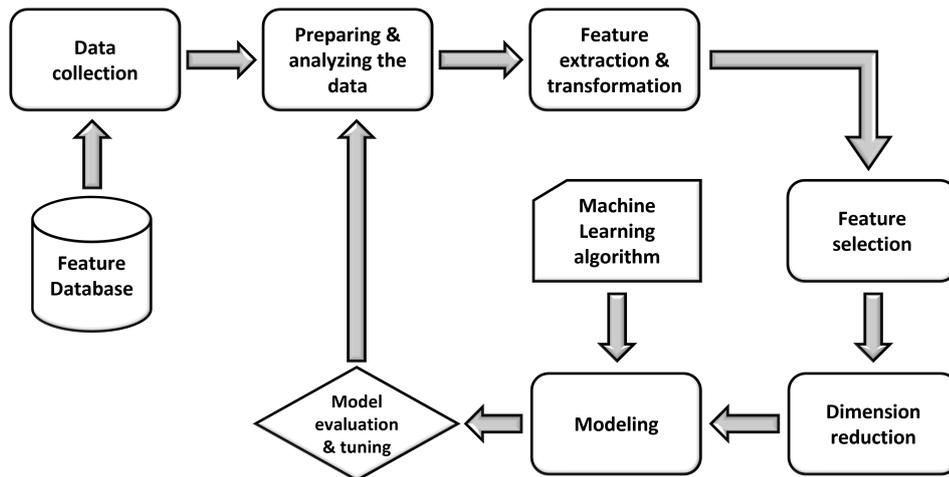


Figura 3.3: Diagramma di flusso delle operazioni caratterizzanti un contesto ML.

3.3.1 Raccolta e descrizione dei dati

Questa fase consiste nell'estrarre, registrare e raccogliere in un *database* i dati necessari all'analisi che si intende svolgere. In base al fenomeno che si intende investigare, talvolta è possibile attingere anche a pre-esistenti banche dati riducendo il tempo e la fatica richiesti in questo *step* preliminare e allo stesso tempo fondamentale del processo. Terminata la fase di raccolta dei dati, è necessario proseguire con la fase di descrizione del dataset. Tale *step* consiste in un'analisi preliminare del dataset che permetta di identificarne la fonte, gli attributi, il volume e le potenziali relazioni tra variabili. Al fine di descrivere opportunamente un dataset, risulta pertanto necessario specificare i seguenti aspetti:

- fonte, identificativa della campagna sperimentale condotta o della banca dati consultata;
- volume di dati, identificativo del numero complessivo di osservazioni registrate;
- attributi, identificativi del numero e della tipologia di variabili registrate;
- relazioni, identificative delle potenziali correlazioni tra diverse variabili;
- descrizione statistica, identificativa dei valori statisticamente significativi come minimo, massimo, media, varianza, ecc.

3.3.2 Analisi esplorativa

Successivamente alla fase descrittiva, l'obiettivo è quello di esplorare e comprendere nel dettaglio il set di dati. Tutte le principali tecniche di visualizzazione come tabelle, grafici e diagrammi diventano utili nell'esaminare i diversi attributi ed identificare le potenziali relazioni e correlazioni. Le variabili e i loro attributi possono essere i più disparati, come stringhe, numeri, ecc. Inoltre, alcuni algoritmi richiedono che il dataset possieda un formato specifico. Pertanto, è opportuno assicurarsi che i dati raccolti siano effettivamente utilizzabili. In questa fase, i dati subiscono un'iniziale elaborazione (anche definita come *pre-processing*) che consiste nel trattamento ed in un'opportuna pulizia degli stessi. I principali compiti svolti durante la fase di analisi esplorativa dei dati possono essere così sintetizzati:

- analizzare, descrivere e visualizzare gli attributi;
- individuare sottoinsiemi di dati e attributi ritenuti cruciali per il problema in esame;
- valutare e verificare la presenza di relazioni ed associazioni;
- rilevare la presenza di eventuali dati mancanti.

3.3.3 Analisi qualitativa

Con l'analisi qualitativa ha inizio la vera e propria comprensione del dataset in esame. Si determinano potenziali dati mancanti, errori o problemi che potrebbero inficiare le operazioni successive. Infatti, tali problematiche necessitano di essere risolte prima di addentrarsi in analisi più approfondite e nella creazione di modelli di apprendimento automatico. I dati devono essere controllati per individuare eventuali schemi all'interno del dataset oppure evidenziare la presenza di dati eccessivamente diversi dagli altri (*outliers*). A tale scopo, può risultare estremamente utile diagrammare i dati in diverse dimensioni. In definitiva, l'analisi qualitativa dei dati focalizza l'attenzione su:

- dati mancanti;
- dati inconsistenti;
- errori (sia lato uomo che lato macchina) che producono informazioni errate.

3.3.4 Preparazione ed integrazione

Dopo aver acquisito una sufficiente comprensione del problema in esame e del relativo set di dati, ha inizio la fase di preparazione dei dati. Questa consiste per lo più in un insieme di attività volte a pulire, organizzare e preparare i dati prima che vengano eseguiti gli algoritmi di apprendimento automatico per la creazione dei modelli. Sebbene tale fase possa risultare lunga e dispendiosa, è necessario che venga svolta con la massima attenzione. Infatti, a prescindere dal singolo algoritmo più o meno sofisticato di apprendimento automatico, un dataset scadente conduce inevitabilmente a modelli scadenti ed a risultati inadeguati [Sarkar et al., 2018]. Qualora si abbiano a disposizione molteplici insiemi di dati relativi allo stesso fenomeno in esame, potrebbe

risultare necessario unirli o quantomeno integrarli. Ciò può avvenire sostanzialmente in due modi:

- se gli attributi risultano inalterati al variare degli insiemi di dati, allora sarà possibile combinarli semplicemente aggiungendo le osservazioni di un dataset in coda a quelle dell'altro;
- se alcuni attributi cambiano al variare degli insiemi di dati, allora sarà possibile unirli mantenendo però solamente gli attributi in comune (che prenderanno il nome di *key features* [Sarkar et al., 2018]).

3.3.5 Gestione e normalizzazione

I dati nella loro forma "grezza" sono difficilmente utilizzabili dagli algoritmi di apprendimento automatico per la creazione dei modelli. Pertanto, tali dati vanno obbligatoriamente gestiti. La gestione dei dati prevede dunque tutte quelle operazioni di elaborazione e formattazione necessarie al fine di un corretto funzionamento degli algoritmi. Essa si focalizza sulla:

- gestione dei valori non chiari degli attributi;
- formattazione dei dati nei formati necessari (ad esempio CSV, JSON, ecc.)

Inoltre, la normalizzazione rappresenta una fase essenziale precedente alla modellazione che comporta la trasformazione dei valori di ciascuna variabile in un intervallo comune, in modo che i valori maggiori da un punto di vista meramente numerico non possano dominare su quelli minori. Lo scopo principale di questa tecnica è quello di ridurre al minimo il Bias legato all'errata associazione tra caratteristiche con contributi numerici maggiori ed effettiva importanza della caratteristica in esame. Dal momento che l'importanza relativa di ciascuna variabile è inizialmente sconosciuta, l'utilizzo di questa pratica consente di partire da una situazione in cui tutte le variabili contribuiscano in maniera eguale al processo di apprendimento automatico e rivestano la stessa importanza nella predizione dell'output desiderato [García et al., 2015].

3.3.6 Estrazione e selezione delle caratteristiche

In molti casi, non è necessario che tutte le variabili registrate concorrano alla realizzazione del modello di apprendimento automatico. Risulta dunque necessario scegliere un sottoinsieme degli attributi esistenti in base alla loro qualità, alla loro importanza ed al loro significato. Tale procedura prende il nome di selezione delle caratteristiche. Talvolta può risultare utile produrre delle nuove variabili a partire da quelle registrate sulla base di determinate logiche, regole o ipotesi. In tal caso si parla di estrazione delle caratteristiche. In questo modo è possibile ridurre la dimensione del vettore delle caratteristiche conservando le informazioni più importanti e rappresentative, rimuovendo quelle meno utili e non correlate. Esistono molteplici tecniche implementabili a tali scopi ed è stato dimostrato che i risultati registrati migliorano di gran lunga, a fronte di ridotti tempi computazionali [Phinyomark et al., 2013, Wołczowski and Zdunek, 2017].

3.3.7 Modellazione

Nella fase di modellazione, i dati e le osservazioni vengono forniti ad un algoritmo di apprendimento automatico affinché un modello possa essere realizzato (Figura 3.4). Migliori sono i dati in termini di pulizia e formattazione, maggiore sarà la probabilità di successo del modello messo a punto. Naturalmente, l'algoritmo di apprendimento automatico dovrà essere poi ottimizzato al fine di ridurre gli errori ed aumentare le sue capacità di generalizzazione. Come rappresentato in Figura 3.3, si tratta di un processo iterativo in cui l'idea di base è quella di costruire una vasta molteplicità di modelli e di addestrarli affinché possa essere identificato quello ottimale in termini di:

- prestazioni definite sulla base di criteri di successo;
- riproducibilità dei risultati e affidabilità;
- robustezza e semplicità di implementazione;
- eventuali assunzioni e/o limitazioni;
- complessità dell'intero sistema di apprendimento automatico, dall'estrazione ed elaborazione dei dati alla modellazione predittiva.

Il punto chiave è tenere traccia dei modelli prodotti, dei parametri utilizzati e dei risultati ottenuti [Sarkar et al., 2018]. È compito dell'ingegnere che si occupa dell'analisi dei dati individuare le tecniche più robuste ed adatte a risolvere il problema in esame. In tal senso queste vengono scelte principalmente sulla base dei dati disponibili, degli obiettivi prefissati, dei vincoli e delle limitazioni proprie di ciascun algoritmo.

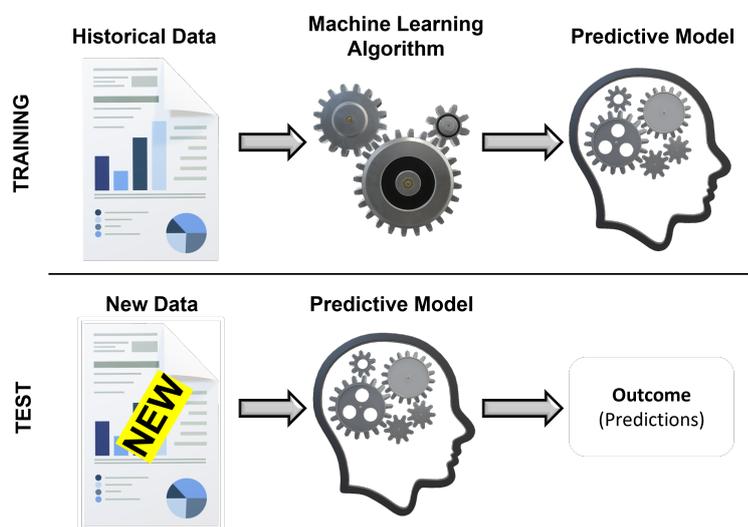


Figura 3.4: Struttura generale di un modello predittivo di apprendimento automatico.

3.4 Compromesso Bias-Varianza

Un concetto chiave nell'apprendimento automatico è rappresentato dal cosiddetto compromesso Bias-Varianza. Tale concetto fa riferimento alla comprensione degli errori di predizione (Bias e varianza) e più in generale ad un equilibrio tra accuratezza e coerenza, risultato del compromesso tra le capacità di un modello di minimizzare le due suddette grandezze. Una corretta comprensione di questi errori aiuta non solo a costruire modelli accurati, ma anche ad evitare spiacevoli fenomeni di underfitting ed overfitting.

3.4.1 Cosa si intende per Bias

Con il termine Bias si definisce la differenza tra la previsione media di un modello ed il valore corretto che si tenta di prevedere [Hellström et al., 2020]. Un modello risulta caratterizzato da un elevato Bias quando è eccessivamente semplificato e presta poca attenzione ai dati di addestramento. Questo conduce ad errori elevati commessi sia sui dati di addestramento che su quelli di test.

3.4.2 Cosa si intende per Varianza

Con il termine varianza si definisce la variabilità della predizione di un modello a partire da una data osservazione. Un modello caratterizzato da un'elevata varianza presta molta attenzione ai dati di addestramento e, pertanto, non è in grado di generalizzare correttamente un set di dati che non ha ancora avuto modo di visualizzare. Di conseguenza, tale modello restituirà delle performance ottime in riferimento ai dati di addestramento ma mostrerà tassi di errore elevati in riferimento ai dati di test [Bouthillier et al., 2021].

3.4.3 Spiegazione analitica

Ipotizzando di indicare con Y la variabile da predire e con X il set di variabili di input, si assume che esista una relazione del tipo:

$$Y = f(X) + \epsilon \quad (3.4.1)$$

dove ϵ sta ad indicare l'errore con distribuzione normale e media nulla. Tramite l'apprendimento automatico si cercherà di mettere a punto un modello $\hat{f}(X)$ di $f(X)$ utilizzando un qualsivoglia algoritmo. Di conseguenza, assegnata un'osservazione x , l'errore quadratico medio (Mean Squared Error - MSE) potrà essere espresso mediante la formula:

$$\begin{aligned} MSE &\triangleq E[(Y - \hat{f}(x))^2] \\ &= E[Y^2 - 2Y\hat{f}(x) + \hat{f}^2(x)] \\ &= E[Y^2] - 2E[Y\hat{f}(x)] + E[\hat{f}^2(x)] \end{aligned} \quad (3.4.2)$$

Dal momento che:

$$Var[x] \triangleq E[(x - E[x])^2] = E[x^2] - E[x]^2 \quad (3.4.3)$$

risulta:

$$E[\hat{f}^2(x)] = Var(\hat{f}) + E[\hat{f}]^2 \quad (3.4.4)$$

In secondo luogo, considerando l'equazione 3.4.1 e semplificando le notazioni $f(x)$ e $\hat{f}(x)$ in f e \hat{f} , risulta che:

$$\begin{aligned} E[Y^2] &= E[(f + \epsilon)^2] \\ &= E[f^2] + 2E[f\epsilon] + E[\epsilon^2] \\ &= f^2 + 2fE[\epsilon] + E[\epsilon^2] \\ &= f^2 + 2f \cdot 0 + \sigma_\epsilon^2 \end{aligned} \quad (3.4.5)$$

Inoltre:

$$\begin{aligned} E[Y\hat{f}] &= E[(f + \epsilon)\hat{f}] \\ &= E[f\hat{f}] + E[\epsilon\hat{f}] \\ &= E[f\hat{f}] + E[\epsilon]E[\hat{f}] \\ &= fE[\hat{f}] \end{aligned} \quad (3.4.6)$$

Infine, combinando le equazioni 3.4.4, 3.4.5 e 3.4.6 si ottiene che:

$$\begin{aligned} MSE &= f^2 + \sigma_\epsilon^2 - 2fE[\hat{f}] + Var[\hat{f}] + E[\hat{f}]^2 \\ &= (f - E[\hat{f}])^2 + Var[\hat{f}] + \sigma_\epsilon^2 \\ &= Bias[\hat{f}]^2 + Var[\hat{f}] + \sigma_\epsilon^2 \\ &= Bias^2 + Variance + Irreducible Error \end{aligned} \quad (3.4.7)$$

L'errore quadratico medio risulta essere pari alla somma del quadrato del Bias, della varianza e di un errore cosiddetto irriducibile. Dunque è importante tenere a mente che, indipendentemente dalla qualità del modello sviluppato, i dati presenteranno una certa quantità di rumore (o di errore irriducibile) che non può essere eliminata. In altre parole, tale frazione irriducibile dell'errore rappresenta una quantità che non può essere ridotta a prescindere dalla bontà dei modelli. Una rappresentazione grafica dei significati di Bias e Varianza viene fornita dal cosiddetto diagramma a bersaglio (conosciuto anche come *Bulls-eye diagram*) riportato in Figura 3.5. Il centro del bersaglio rappresenta una predizione corretta. Man mano che tale predizione si allontana dal centro peggiora e diventa via via sempre meno attendibile.

3.4.4 Fenomeni di Underfitting ed Overfitting

Nell'ambito dell'apprendimento automatico supervisionato, è pratica ormai consolidata valutare i modelli in base alle loro prestazioni ed ai livelli di accuratezza raggiunti. In linea generale, un modello si definisce ben realizzato se è in grado di generalizzare correttamente qualsiasi nuovo set di dati gli venga fornito in ingresso, a patto che rientri nel dominio del problema in esame. Questo permette di effettuare delle previsioni più o meno accurate sulla base di dati

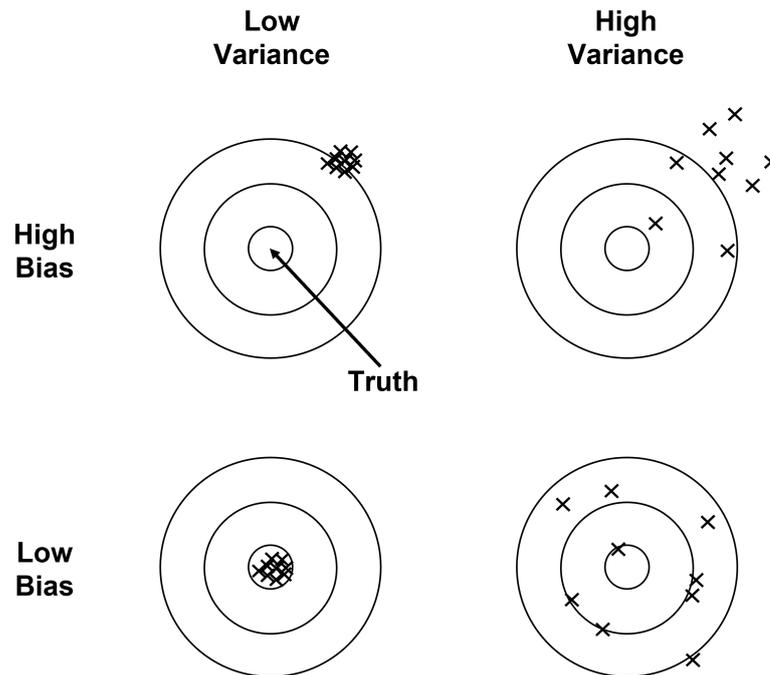


Figura 3.5: Diagramma a bersaglio.

che il modello non ha ancora avuto modo di visualizzare. I principali fenomeni responsabili delle scarse prestazioni degli algoritmi di apprendimento automatico supervisionato prendono il nome di sotto-adattamento (*underfitting*) e sovra-adattamento (*overfitting*).

Un modello si definisce *underfitted* quando non è in grado di catturare la tendenza di fondo dei dati che sta analizzando [Bashir et al., 2020]. Pertanto, l'algoritmo messo a punto non riesce ad adattarsi sufficientemente bene ai dati in esame. Le principali cause del fenomeno di *underfitting* sono le seguenti:

- le dimensioni del set di dati utilizzato per costruire il modello sono eccessivamente ridotte e non sufficienti;
- il set di dati utilizzato per addestrare il modello non è stato opportunamente depurato dei valori anomali e pertanto contiene un'eccessiva quantità di "rumore";
- si cerca di modellare un fenomeno complesso e/o non lineare sulla base di un modello semplice e/o lineare;
- il numero delle epoche di addestramento è eccessivamente ridotto.

La presenza di uno o più fattori tra quelli appena elencati provoca inevitabilmente che il modello faccia delle previsioni errate, caratterizzate da una ridotta varianza e da un elevato Bias.

Al contrario, un modello si definisce *overfitted* quando, sebbene le prestazioni raggiunte in fase di addestramento siano sufficientemente valide, le previsioni sui dati di test risultano poco

accurate [Ying, 2019]. Tale fenomeno si verifica quando un modello apprende i dettagli ed il "rumore" dei dati di addestramento in misura tale da influire negativamente sulle sue prestazioni su nuovi dati (ad esempio di test). Ciò significa che il rumore o le fluttuazioni casuali nei dati di addestramento vengono raccolti e appresi dal modello come concetti. Il problema è che questi concetti non si applicano ai nuovi dati e influiscono negativamente sulle capacità di generalizzazione del modello, risultando in previsioni caratterizzate da Bias ridotto ed elevata varianza. Pertanto, un algoritmo ben funzionante dovrebbe essere in grado di distinguere i dati rappresentativi dai rumori [Paris et al., 2003]. Le principali cause del fenomeno di *overfitting* sono le seguenti:

- il set di addestramento ha dimensioni troppo limitate;
- il set di addestramento è costituito da dati poco rappresentativi;
- il set di addestramento presenta un eccessivo rumore nei dati;
- eccessiva complessità delle ipotesi. Quando gli algoritmi presentano un numero elevato di ipotesi (ovvero possiedono un'elevata quantità di input), il modello diventa mediamente più accurato ma con una minore coerenza [Paris et al., 2003].

Riassumendo, un modello troppo semplice caratterizzato da un eccessivamente ridotto numero di parametri restituisce Bias elevato e varianza ridotta (*underfitting*). Tuttavia, un modello troppo complesso caratterizzato da un eccessivo numero di parametri restituisce Bias ridotto e varianza elevata (*overfitting*). Lo scopo della modellazione è quindi di identificare il giusto compromesso tra questi due fenomeni in modo da minimizzare l'errore complessivo (Figura 3.6).

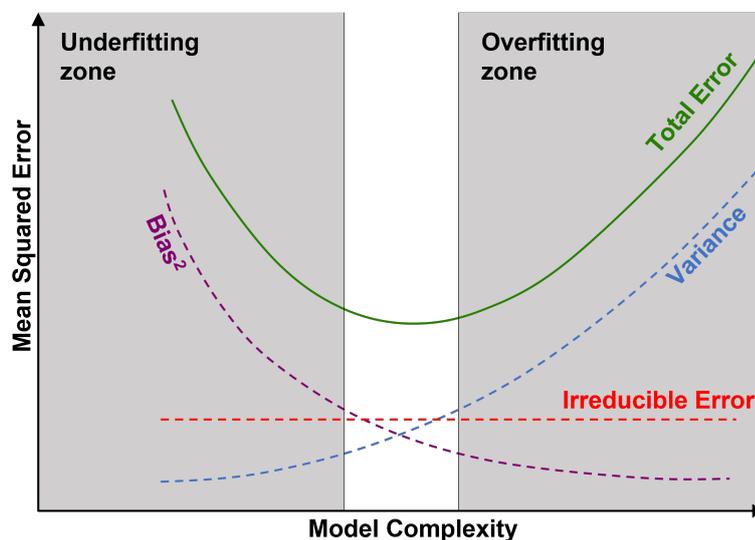


Figura 3.6: Rappresentazione grafica del compromesso Bias-varianza.

3.5 Early Stopping

Uno degli espedienti più comunemente utilizzati per migliorare le capacità di generalizzazione dei modelli evitando che si verifichino fenomeni di *overfitting* prende il nome di arresto anticipato (*early stopping*). Consiste nell'utilizzare un insieme di dati di validazione (ad esempio estratto a partire dal set di addestramento) per valutare e monitorare l'andamento dell'errore. Via via che l'addestramento prosegue, è lecito immaginare che l'errore commesso sul set di addestramento diminuisca, mentre l'errore commesso sul set di validazione avrà una fase di decrescita seguita da una fase di crescita. Si definisce punto di arresto anticipato, quello in cui l'errore sul set di validazione raggiunge il minimo. In tale punto, il modello raggiunge le migliori capacità di generalizzazione [Prechelt, 2002].

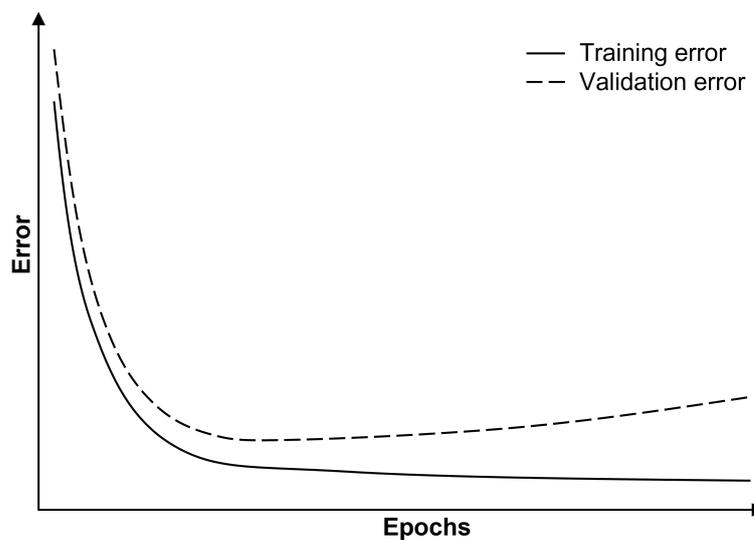


Figura 3.7: Curve ideali di errore in fase di addestramento e validazione.

Sulla base del comportamento evidenziato in Figura 3.7, l'applicazione della tecnica di *early stopping* consiste nei seguenti *step* procedurali:

1. suddividere l'insieme di dati di addestramento in due insiemi, uno di addestramento ed uno di validazione in proporzione 2 a 1 oppure 3 a 1;
2. eseguire l'addestramento sull'insieme di dati preposto all'addestramento e saltuariamente valutare l'errore commesso sull'insieme di dati di validazione, ad esempio ogni cinque epoche;
3. interrompere l'addestramento qualora l'errore commesso sull'insieme di addestramento sia superiore rispetto a quello precedentemente registrato;
4. utilizzare i parametri registrati nella fase precedente come risultato finale dell'addestramento.

In tal senso, questo approccio utilizza il set di validazione per anticipare il comportamento che il modello avrebbe in un utilizzo reale (oppure su un set di test), assumendo che l'errore commesso sarà simile. Pertanto, l'errore di validazione viene utilizzato come stima dell'errore di generalizzazione. Le reali curve relative agli errori di validazione tuttavia non evolvono sempre in maniera così regolare e presentano spesso più di un minimo locale. Non esiste quindi un metodo univoco per identificare il punto in cui si ottiene il minimo globale dell'errore di generalizzazione. C'è bisogno di una regola che indichi quando interrompere l'addestramento e tale regola prende il nome di criterio d'arresto. La scelta di un criterio d'arresto comporta prevalentemente un compromesso tra tempo computazionale di addestramento del modello ed errore di generalizzazione. Tra tutti i possibili criteri d'arresto si cerca quindi quello che permetterà al modello di raggiungere un predeterminato errore di generalizzazione nel tempo più breve possibile oppure l'errore di generalizzazione più basso a fronte di un assegnato tempo di addestramento.

3.5.1 Criteri d'arresto

Esistono diversi potenziali criteri d'arresto e, nel seguito, verranno trattate solamente le tre classi principali. Per descrivere formalmente i criteri d'arresto sono necessarie alcune definizioni. Sia E la funzione obiettivo, o funzione d'errore, dell'algoritmo di addestramento (ad esempio l'errore quadratico medio). Siano $E_{tr}(t)$, $E_{va}(t)$ e $E_{te}(t)$ gli errori commessi sul set di addestramento, di validazione e di test rispettivamente misurati all'epoca t . Il valore ottimale $E_{opt}(t)$ definisce l'errore di validazione più basso ottenuto nelle epoche precedenti t :

$$E_{opt}(t) := \min_{t' \leq t} E_{va}(t') \quad (3.5.1)$$

Si definisce perdita di generalizzazione (GL) all'epoca t l'incremento relativo dell'errore di validazione in rapporto al valore ottimale, espresso in percentuale:

$$GL(t) = 100 \cdot \left(\frac{E_{va}(t)}{E_{opt}(t)} - 1 \right) \quad (3.5.2)$$

Un valore elevato di perdita di generalizzazione è un valido motivo per interrompere l'addestramento poiché indica la presenza di *overfitting*. Questa considerazione conduce alla definizione della prima classe dei criteri di arresto, ovvero quella classe che interrompe l'addestramento qualora la perdita di generalizzazione superi una certa soglia α .

Si definisce la classe GL_α come quella classe che interrompe l'addestramento dopo l'epoca t che registra:

$$GL_\alpha : GL(t) > \alpha \quad (3.5.3)$$

Tutt'altra considerazione si trova alla base della seconda classe dei criteri d'arresto. Si può infatti pensare di voler interrompere l'addestramento in un periodo in cui l'errore si trova in fase di diminuzione rapida. Questo perché il fenomeno di *overfitting* si verifica quando l'errore inizia a diminuire in maniera lenta. Pertanto, è lecito pensare che, in una porzione del tratto fortemente discendente della curva dell'errore di addestramento, le capacità di generalizzazione

del modello siano sufficientemente valide. Al fine di formalizzare tale considerazione, si consideri una striscia di epoche di addestramento di lunghezza k (numerata come $n + 1, n + 2, \dots, n + k$ con n divisibile per k). Si definisce progresso di addestramento (per mille) misurato dopo tale striscia:

$$P_k(t) := 1000 \cdot \left(\frac{\sum_{t'=t-k+1}^t E_{tr}(t')}{k \cdot \min_{t'=t-k+1}^t E_{tr}(t')} - 1 \right) \quad (3.5.4)$$

$P_k(t)$ risponde al quesito: "di quanto la media dell'errore di addestramento all'interno della striscia k era maggiore dell'errore minimo di addestramento commesso all'interno della stessa striscia?". Tale definizione consente dunque di definire la seconda classe dei criteri di arresto, ovvero quella classe che interrompe l'addestramento qualora il rapporto tra la perdita di generalizzazione ed il progresso di addestramento superi una certa soglia α .

Si definisce la classe PQ_α come quella classe che interrompe l'addestramento al termine della striscia con epoca t che registra:

$$PQ_\alpha : \frac{GL(t)}{P_k(t)} > \alpha \quad (3.5.5)$$

Solitamente si assume una lunghezza delle strisce pari a 5 epoche e si misura l'errore di validazione esclusivamente al termine di ciascuna striscia [Prechelt, 2002].

La terza classe dei criteri di arresto si basa sul segno delle variazioni dell'errore di generalizzazione, ovvero quella classe che interrompe l'addestramento qualora l'errore di generalizzazione incrementi in s strisce consecutivamente successive. L'idea alla base di questa definizione scaturisce dalla considerazione per cui, se l'errore aumenta per s strisce consecutive, tali aumenti indichino un principio del fenomeno di *overfitting* (indipendentemente da quanto siano effettivamente elevati). In definitiva, si definisce la classe UP_s come quella classe che interrompe l'addestramento al verificarsi per un numero fissato di strisce s della seguente condizione:

$$UP_s : E_{va}(t) > E_{va}(t - k) \quad (3.5.6)$$

Tutti i sopracitati criteri d'arresto svolgono lo stesso ruolo all'interno dello sviluppo di un modello di apprendimento automatico. Essi determinano, infatti, a che punto interrompere l'addestramento e quale insieme di parametri fissare affinché l'errore risultante in fase di validazione sia il più basso possibile ed uguale a $E_{opt}(t)$. In linea generale, è stato evidenziato che criteri d'arresto più lenti (ovvero che interrompono l'addestramento più tardi rispetto agli altri) mediamente conducono a capacità di generalizzazione superiori rispetto ai criteri più veloci [Fahlman et al., 1988]. Tuttavia, nonostante le differenze sistematiche tra l'applicazione di un criterio piuttosto che un altro siano relativamente piccole, le seguenti regole empiriche sono state riscontrate:

- al fine di massimizzare la probabilità di identificare una singola buona soluzione (in contrapposizione al massimizzare la qualità media delle soluzioni), è preferibile utilizzare un criterio GL_α ;
- al fine di massimizzare la qualità media delle soluzioni, è preferibile utilizzare un criterio PQ_α se il modello è solo leggermente sovradattato, altrimenti un criterio UP_s .

3.6 Convalida Incrociata

Assegnato un insieme di dati di partenza, l'utilizzo di un particolare metodo di apprendimento automatico risulta giustificato se conduce ad un errore ridotto in fase di test. Tale errore può essere facilmente calcolato se si dispone di un designato insieme di osservazioni di test. Purtroppo, la condizione appena descritta non si verifica molto spesso. Più frequentemente, infatti, il set di osservazioni di partenza viene diviso in due sottoinsiemi in maniera casuale, addestrando il modello sul primo dei sottoinsiemi e testandolo sul secondo. Tuttavia, l'errore commesso in fase di addestramento solitamente differisce di gran lunga dall'errore commesso in fase di test. Una strategia utile per sopperire a tale problema consiste, come anticipato nel corso del paragrafo 3.5, nel ricavare un ulteriore sottoinsieme di osservazioni che prende il nome di set di validazione. Il modello si addestra dunque sul set di addestramento ed, una volta addestrato, viene utilizzato per effettuare delle previsioni a partire dalle osservazioni presenti nel sottoinsieme di validazione. Gli errori risultanti, tipicamente determinati nella forma dell'errore quadratico medio (MSE), forniscono una stima dei successivi errori in fase di test che verosimilmente, grazie all'utilizzo di tecniche di arresto anticipato ed all'aumentata capacità di generalizzazione, risulteranno più ridotti. Al fine di comprendere più approfonditamente le prestazioni di un modello di apprendimento automatico, i metodi di ricampionamento rappresentano uno strumento indispensabile nelle metodologie più moderne ed innovative. Essi consistono nel prelevare ripetutamente campioni dall'insieme dei dati di addestramento e nel riadattare il modello di interesse su ciascun campione, al fine di ottenere delle informazioni aggiuntive. Chiaramente, tali informazioni non sarebbero state disponibili applicando semplicemente il modello una singola volta su tutto l'insieme originale dei dati di addestramento. I metodi di ricampionamento possono risultare particolarmente onerosi dal punto di vista computazionale dal momento che comportano l'applicazione dello stesso algoritmo più volte, utilizzando ciascuna volta sottoinsiemi diversi dei dati di addestramento. Tuttavia, grazie ai recenti progressi tecnologici relativi alla potenza di calcolo, i requisiti computazionali dei metodi di ricampionamento generalmente non risultano più proibitivi. Nei prossimi paragrafi, verrà analizzata nel dettaglio la pratica di ricampionamento più comunemente utilizzata, la convalida incrociata (o *cross-validation*).

3.6.1 Leave-One-Out Cross-Validation

La convalida incrociata *leave-one-out* (o *Leave-One-Out Cross-Validation* - LOOCV) rappresenta una tecnica di ricampionamento strettamente correlata all'approccio del set di validazione descritto nel paragrafo precedente. Allo stesso modo, infatti, l'approccio LOOCV prevede che l'insieme dei dati di addestramento venga suddiviso in due sottoinsiemi. Tuttavia, in questo caso, i due sottoinsiemi non possiederanno dimensioni confrontabili bensì il set di validazione sarà costituito esclusivamente da una singola osservazione che, per semplicità, viene qui indicata come (x_1, y_1) . Tutte le rimanenti osservazioni, indicate come $\{(x_2, y_2), \dots, (x_n, y_n)\}$ costituiranno il set di addestramento. Il modello di apprendimento automatico verrà dunque addestrato sulla base delle osservazioni che compongono il set di addestramento ed effettuerà, sulla base delle osservazioni x_1 , una previsione del valore di y_1 , indicata come \hat{y}_1 . Dal momento che l'osservazione (x_1, y_1) è stata volutamente esclusa dalla fase di addestramento, è

lecito pensare che l'errore quadratico medio $MSE_1 = (y_1 - \hat{y}_1)^2$ fornisca una stima approssimativamente imparziale dell'errore di test. Tuttavia, sebbene MSE_1 rappresenti una stima imparziale dell'errore di test, tale stima non potrà essere ritenuta sufficiente poiché altamente variabile e legata ad una singola osservazione. È però possibile ripetere la procedura appena descritta selezionando l'osservazione (x_2, y_2) come set di validazione, addestrando il modello di apprendimento automatico sulle rimanenti $n - 1$ osservazioni $\{(x_1, y_1)(x_3, y_3), \dots, (x_n, y_n)\}$ e calcolando $MSE_2 = (y_2 - \hat{y}_2)^2$. Ipotizzando di ripetere n volte questa procedura, è possibile ottenere n diversi valori di MSE ($MSE_1, MSE_2, \dots, MSE_n$). In questo modo, la stima dell'errore di test risulterà pari alla media delle n stime effettuate:

$$CV_{(n)} = \frac{1}{n} \sum_{i=1}^n MSE_i \tag{3.6.1}$$

La Tabella 3.3 illustra uno schema rappresentativo dell'approccio LOOCV. Esso presenta dei notevoli vantaggi rispetto al semplice approccio del set di validazione. In primo luogo, presenta una Bias decisamente ridotto. Questo perché si è andati ad adattare ripetutamente il modello di apprendimento automatico, utilizzando insiemi di addestramento contenenti $n - 1$ osservazioni (ovvero quasi quante quelle contenute nell'intero set di addestramento). In tal senso, l'approccio LOOCV tende a non sovrastimare l'errore di test, a differenza di quanto accade nel semplice approccio del set di validazione in cui tale insieme ha dimensioni circa pari ad un terzo oppure ad un mezzo dell'insieme di addestramento. In secondo luogo, un semplice approccio del set di validazione restituirà risultati diversi se applicato ripetutamente (in conseguenza della casualità nella suddivisione delle osservazioni nei due insiemi). Tale casualità si annulla nell'approccio LOOCV che, pertanto, restituirà sempre il medesimo risultato anche per applicazioni ripetute. Tale procedura risulta chiaramente molto onerosa dal punto di vista computazionale, poiché richiede che il singolo modello debba addestrarsi n diverse volte. Inoltre, risulterà tanto più dispendiosa in termini di tempo quanto più aumenteranno il valore di n ed il tempo di un singolo addestramento. Infine, l'approccio LOOCV rappresenta un metodo molto generale che può essere implementato in qualsiasi algoritmo di modellazione predittiva.

Tabella 3.3: Rappresentazione della procedura Leave-One-Out Cross-Validation.

	Observ. 1	Observ. 2	Observ. 3	Observ. ...	Observ. n	
Split 1	Validation	Train	Train	Train	Train	MSE_1
Split 2	Train	Validation	Train	Train	Train	MSE_2
Split 3	Train	Train	Validation	Train	Train	MSE_3
Split ...	Train	Train	Train	Validation	Train	$MSE \dots$
Split n	Train	Train	Train	Train	Validation	MSE_n

3.6.2 k-fold Cross-Validation

Un'alternativa alla tecnica LOOCV è l'approccio comunemente denominato convalida incrociata *k-fold* (o *k-fold Cross-Validation* - *k-fold CV*). Tale approccio prevede la suddivisione casuale

delle osservazioni che compongono il set di addestramento in k gruppi (denominati *fold*) di dimensioni approssimativamente uguali. Il primo di questi gruppi diventa il set di validazione e l'addestramento viene condotto sui rimanenti $k - 1$ gruppi. L'errore quadratico medio MSE_1 viene quindi calcolato sulla base delle osservazioni contenute nel gruppo di validazione. La procedura viene ripetuta iterativamente k volte; ognuna di queste volte, un gruppo diverso di osservazioni costituisce il set di validazione. Vengono quindi effettuate k diverse stime dell'errore di validazione ($MSE_1, MSE_2, \dots, MSE_k$) ed, analogamente a quanto accade nel caso dell'approccio LOOCV, la stima dell'errore di validazione risulterà pari alla media delle k stime effettuate:

$$CV_{(k)} = \frac{1}{k} \sum_{i=1}^k MSE_i \quad (3.6.2)$$

La Tabella 3.4 illustra uno schema rappresentativo dell'approccio k-fold CV. Si intuisce come l'approccio LOOCV sia un caso speciale dell'approccio k-fold CV in cui il parametro k viene impostato pari a n . È pratica comune utilizzare un valore di k pari a 5 oppure pari a 10 [James et al., 2013]. Il vantaggio principale di utilizzare un valore di $k < n$ è quello computazionale. Infatti, tranne che per algoritmi di apprendimento automatico estremamente semplici, la riduzione del valore di k apporta importanti benefici in termini di costi computazionali e relativi tempi di esecuzione. Quindi, senza perdere in generalità della tecnica, l'approccio k-fold CV può essere ugualmente implementato in qualsiasi algoritmo di modellazione predittiva e rappresenta un'importante alternativa all'approccio LOOCV, specialmente in presenza di ampi dataset, con valori elevati di n .

Tabella 3.4: Rappresentazione della procedura k-fold Cross-Validation.

	Fold 1	Fold 2	Fold 3	Fold ...	Fold k	
Split 1	Validation	Train	Train	Train	Train	MSE_1
Split 2	Train	Validation	Train	Train	Train	MSE_2
Split 3	Train	Train	Validation	Train	Train	MSE_3
Split ...	Train	Train	Train	Validation	Train	$MSE \dots$
Split k	Train	Train	Train	Train	Validation	MSE_k

In linea generale è possibile affermare che nell'analisi di dati reali non sia possibile conoscere il vero valore di MSE in fase di test. Pertanto risulta complesso determinare l'accuratezza della stima ottenuta tramite convalida incrociata. L'unica strada percorribile risulta determinare il valore di MSE calcolato sull'insieme di osservazioni di test e confrontarlo con il valore ottenuto calcolandolo sull'insieme di osservazioni di validazione, al fine di valutare l'accuratezza dei risultati della convalida incrociata.

3.6.3 Compromesso Bias-Varianza nella convalida incrociata

Come accennato nel paragrafo precedente, una procedura k-fold CV con $k < n$ apporta un notevole vantaggio computazionale rispetto ad una LOOCV. Inoltre, un ulteriore vantaggio

probabilmente meno ovvio ma potenzialmente più importante risiede nel fatto che la k-fold CV spesso fornisca stime più accurate del valore di MSE in fase di test rispetto alla LOOCV. Tale fenomeno si verifica in virtù del compromesso Bias-Varianza accennato al paragrafo 3.4. Come ormai noto, il semplice approccio del set di validazione comporta una potenziale sovrastima del valore di MSE di test, poiché l'insieme di dati di addestramento contiene una quantità di osservazioni all'incirca dimezzata rispetto alle dimensioni originali. Sulla base di questa considerazione, è intuibile comprendere che le stime del valore di MSE fornite dall'approccio LOOCV risulteranno invece approssimativamente imparziali dal momento che ogni insieme di addestramento conterrà $n - 1$ osservazioni (un numero quindi pressoché uguale al numero di osservazioni contenute nel set originale). In questo senso, la procedura k-fold CV condurrà ad un livello intermedio di Bias dal momento che ogni insieme di addestramento conterrà all'incirca $\frac{(k-1)n}{k}$ osservazioni, un numero certamente inferiore rispetto a quello prodotto dall'approccio LOOCV ma sostanzialmente superiore rispetto a quello prodotto dall'approccio con semplice set di validazione. In definitiva, dal solo punto di vista della riduzione del Bias, è chiaro che l'approccio LOOCV sarebbe da preferire all'approccio k-fold CV. Tuttavia, bisogna tener conto anche della Varianza prodotta dalla procedura. Dalle analisi condotte sulle diverse procedure, si evince che l'approccio LOOCV conduce a valori di varianza maggiori rispetto a quelli ottenuti tramite approccio k-fold CV. Ciò accade perché, eseguendo una LOOCV, si effettua una media degli output prodotti da n modelli addestrati su un insieme quasi identico di osservazioni; pertanto, tali output risultano altamente (positivamente) correlati tra loro. Al contrario, eseguendo una k-fold CV con $k < n$, si effettua una media degli output prodotti da k modelli; tali output risulteranno sicuramente meno correlati tra loro poiché la sovrapposizione tra insiemi di addestramento di ciascun modello è certamente minore. Poiché la media di molteplici quantità altamente correlate ha una varianza più elevata rispetto alla media effettuata su quantità non altrettanto correlate, la stima dell'errore di test risultante dalla LOOCV tende ad avere una varianza più elevata rispetto alla stima dell'errore di test risultante dalla k-fold CV. In sintesi, esiste un compromesso Bias-Varianza associato alla scelta di k : in virtù di questo compromesso, solitamente si esegue una k-fold CV utilizzando valori di k pari a 5 o a 10. Questo perché tali valori hanno dimostrato empiricamente di produrre stime dell'errore di test che non soffrono né di Bias né di Varianza eccessivamente elevati [James et al., 2013].

3.7 Regolazione degli iperparametri

I modelli di apprendimento automatico sono fondamentalmente funzioni matematiche in grado di rappresentare la relazione tra diversi aspetti dei dati. Per esempio, un modello di regressione lineare utilizza una retta per rappresentare la relazione tra caratteristiche ed obiettivo:

$$y = w^T \cdot x \quad (3.7.1)$$

dove x è un vettore che rappresenta le caratteristiche dei dati di input mentre y è una variabile scalare che rappresenta l'obiettivo della modellazione, ovvero una quantità numerica che si vuole imparare a prevedere. Il modello appena descritto presuppone che la relazione tra x e y sia di tipo lineare. La variabile w è il vettore dei pesi e rappresenta il vettore normale alla retta, ossia specifica la pendenza della stessa. Si tratta di un cosiddetto parametro del modello, appreso

durante la fase di addestramento. Pertanto, l'addestramento di un modello comporta l'utilizzo di procedure di ottimizzazione in grado di determinare i migliori parametri del modello adattandosi ai dati a disposizione. Esiste, però, un'altra serie di parametri noti come *iperparametri*. Tali valori, a differenza dei parametri, devono essere specificati al di fuori della procedura di addestramento. La regressione lineare classica non presenta iperparametri. Tuttavia, sono molteplici gli algoritmi di apprendimento automatico che presentano iperparametri in numero molto variabile. Ad esempio: alcune varianti di regressione (regressioni *Ridge* e *Lasso*) presentano i cosiddetti termini di regolarizzazione, gli alberi decisionali presentano iperparametri come la profondità dell'albero ed il numero di foglie, le macchine a vettori di supporto richiedono l'impostazione dei termini di penalizzazione oppure dei parametri del kernel e così via. In linea generale, è possibile definire iperparametri di:

- regolarizzazione, cioè legati all'architettura del modello. Determinano la flessibilità del modello ed i gradi di libertà a sua disposizione per adattarsi ai dati. Il controllo di questo genere di iperparametri può evitare che si verifichino fenomeni di *overfitting*;
- addestramento, cioè legati al funzionamento dell'algoritmo di *training*. A titolo esemplificativo si citano il tasso di apprendimento ed il numero complessivo di iterazioni dell'algoritmo.

Concettualmente, la regolazione degli iperparametri è un problema di ottimizzazione, proprio come l'addestramento del modello. Tuttavia, nella pratica comune questi due compiti si avvicinano in maniera completamente diversa. Durante l'addestramento di un modello, la qualità di un insieme di parametri del modello proposto può essere valutata mediante una formula matematica (di solito chiamata funzione di perdita o di costo). La qualità degli iperparametri, invece, non può essere scritta in formula chiusa, poiché dipende dal risultato di un processo di addestramento. Per questo motivo, la messa a punto degli iperparametri è particolarmente complessa. Fino a qualche anno fa, le uniche metodologie disponibili ad assolvere tale compito erano la ricerca a griglia e quella casuale. Negli ultimi anni, l'interesse per l'*auto-tuning* è aumentato. Diversi gruppi di ricerca si sono occupati (e tutt'ora si occupano) del problema, pubblicando articoli e rilasciando nuovi strumenti utili all'ottimizzazione degli iperparametri.

3.7.1 Ricerca a griglia

La ricerca a griglia (*grid search*), come suggerito dal nome, consiste nel selezionare una griglia di potenziali valori degli iperparametri, valutarne ogni singola combinazione e restituire quella ottimale. Ad esempio, nel valutare l'iperparametro "numero di foglie" di un albero decisionale, la griglia potrebbe essere {10, 20, 30, ..., 100}. Per i parametri di regolarizzazione, invece, è pratica comune utilizzare una scala di ricerca logaritmica come { 10^{-5} , 10^{-4} , 10^{-3} , ..., 1}. Per identificare i valori minimi e massimi della griglia per ciascun iperparametro da ottimizzare è necessario effettuare qualche congettura oppure fare riferimento ai più rilevanti riferimenti bibliografici. Talvolta si esegue la ricerca su una piccola griglia, si verifica se la combinazione ideale si trova in uno dei due punti estremi della griglia e la si espande progressivamente in quella direzione. Tale pratica prende il nome di ricerca manuale della griglia. La ricerca a griglia è la più semplice metodologia di ottimizzazione degli iperparametri da impostare e da

parallelizzare. Pertanto, si tratta del metodo più oneroso in termini di calcolo computazionale; tuttavia, se eseguito in parallelo, può risultare il più veloce in termini di durata complessiva.

3.7.2 Ricerca casuale

La ricerca casuale (*random search*) rappresenta una leggera variazione della sopracitata ricerca a griglia. Infatti, piuttosto che investigare l'intera griglia, la ricerca casuale focalizza le sue valutazioni solo su un campione casuale di punti della griglia. Questo fattore rende la ricerca casuale molto meno onerosa della ricerca a griglia dal punto di vista computazionale. In passato la ricerca casuale non è stata particolarmente apprezzata in ambito scientifico poiché non effettua la ricerca su tutti i punti della griglia e, pertanto, non può battere la combinazione ideale identificata dalla ricerca a griglia. Tuttavia, è stato dimostrato che, in un numero sorprendente di casi, la ricerca casuale si comporta bene tanto quanto quella a griglia a patto che la ricerca venga effettuata su almeno 60 punti casuali campionati dalla griglia [Bergstra and Bengio, 2012]. Tale valore risulta sufficiente in virtù di una spiegazione probabilistica: per qualsiasi distribuzione su uno spazio campionario con un massimo finito, il massimo ottenuto tramite 60 tentativi casuali rientra all'interno del 5% del massimo assoluto, con una probabilità del 95%. Immaginando l'intervallo del 5% intorno al massimo assoluto, è possibile campionare dei punti dalla griglia e valutare se ricadano all'interno di questo intervallo. Ogni estrazione casuale ha una probabilità del 5% di rientrare in questo intervallo; ipotizzando di estrarre n punti in modo indipendente, la probabilità che tutti manchino l'intervallo desiderato è $(1 - 0,05)^n$. Pertanto, la probabilità che almeno uno di essi rientri nell'intervallo è pari a $1 - (1 - 0,05)^n$. La probabilità obiettivo dev'essere almeno pari al 95%. Per calcolare il numero di estrazioni necessarie, basta risolvere per n la seguente disequazione: $1 - (1 - 0,05)^n \geq 0,95$. Il risultato è proprio $n \geq 60$. Pertanto, se la regione quasi-ottima degli iperparametri occupa almeno il 5% della superficie della griglia, una ricerca casuale con almeno 60 tentativi identificherà quella regione con una probabilità maggiore del 95%. L'assunzione iniziale è molto importante e può essere soddisfatta se la regione quasi-ottimale è grande o se in qualche modo c'è un'alta concentrazione di punti della griglia in quella regione. La prima ipotesi è quella più probabile, poiché un buon modello di apprendimento automatico non dovrebbe essere eccessivamente sensibile agli iperparametri. Pertanto, la ricerca casuale richiede un numero molto inferiore di tentativi rispetto alla ricerca a griglia, è ugualmente parallelizzabile e presenta prestazioni confrontabili nella maggior parte dei casi.

3.7.3 Ricerca intelligente

A differenza delle alternative più semplici come la ricerca a griglia o quella casuale, le tecniche di ricerca intelligente scelgono alcune impostazioni degli iperparametri, valutano la qualità del modello corrispondente e decidono in maniera più o meno autonoma quale zona dello spazio di ricerca campionare durante il tentativo successivo. Si tratta di processi meno parallelizzabili ma intrinsecamente iterativi e sequenziali, atti a ridurre il numero di valutazioni ed i corrispondenti tempi computazionali. La scelta "intelligente" del successivo punto di campionamento richiede del tempo. Pertanto, l'utilizzo di queste tecniche risulta tanto più utile quanto più lungo e complesso è il processo di addestramento e valutazione del modello cosiddetto "interno" rispetto

alla procedura "esterna" di selezione del set di iperparametri. Tra i metodi di ottimizzazione intelligente degli iperparametri, a titolo esemplificativo si citano: i metodi senza derivate (*derivative-free*), l'ottimizzazione con foresta casuale (*random forest*) e l'ottimizzazione Bayesiana (*Bayesian Optimization*).

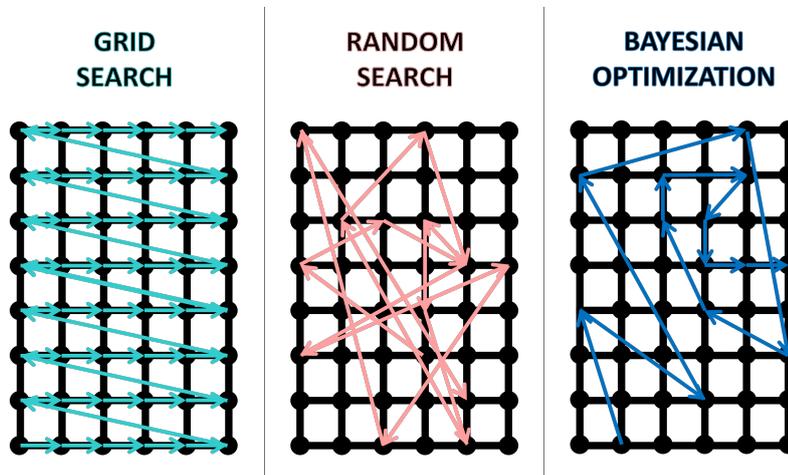


Figura 3.8: Tre degli algoritmi più diffusi per l'ottimizzazione degli iperparametri.

3.7.4 Ottimizzazione Bayesiana

L'ottimizzazione Bayesiana degli iperparametri rappresenta un metodo basato sulla statistica Bayesiana che ha lo scopo di identificare il massimo (nei problemi di massimizzazione) o, alternativamente, il minimo (nei problemi di minimizzazione) di una funzione $f(\mathbf{x})$, $\mathbf{x} = [x_p]$, $p \in \{0, \dots, P\}$, $P \in \mathbb{N}$ in cui x_p rappresenta un parametro definito all'interno di un dominio limitato $X_p \subset \mathbb{R}$. Tale problema matematico può essere risolto attraverso algoritmi di ottimizzazione che definiscono un modello probabilistico di $f(x)$ per decidere, durante ciascuna iterazione, quale sia il punto più probabile di X in grado di massimizzare (o minimizzare) $f(\cdot)$. In tale contesto, nel 2012 [Snoek et al., 2012] sono stati i primi ad utilizzare l'ottimizzazione Bayesiana per la ricerca dei migliori iperparametri di un modello ML. Poiché l'andamento della funzione obiettivo è inizialmente sconosciuto, gli autori hanno trattato $f(\cdot)$ come una funzione casuale e l'hanno approssimata a priori con un processo gaussiano (*Gaussian Process - GP*) [Rasmussen, 2003] al fine di catturarne il comportamento. Durante il processo di ottimizzazione, GP viene aggiornato in base ai risultati delle simulazioni ML prodotte da diverse combinazioni degli iperparametri per definire la distribuzione della funzione $f(\cdot)$ a posteriori. Quest'ultima, a sua volta, viene sfruttata da una cosiddetta funzione di acquisizione per determinare il successivo punto di valutazione. Una rappresentazione schematica dell'intera procedura è stata riportata in Figura 3.9.

In pratica, un insieme di osservazioni O nella forma $\{\mathbf{x}^{(o)}, \mathbf{y}^{(o)}\}_{o=1}^O$, con $\mathbf{y}^{(o)} = N(f(\mathbf{x}^{(o)}), v)$ e v la varianza del rumore nell'osservazione di $f(\cdot)$, viene utilizzato per individuare una distribuzione gaussiana multi-variata in \mathbb{R}^O attraverso le funzioni media e covarianza definite

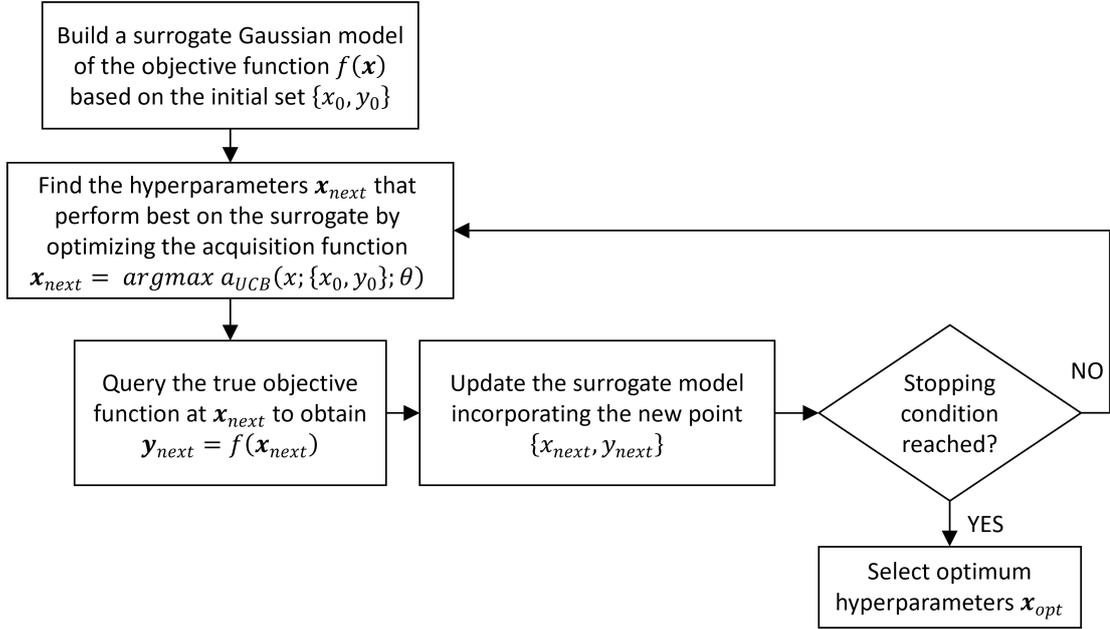


Figura 3.9: Schematizzazione dell'ottimizzazione Bayesiana degli iperparametri.

rispettivamente da $m : X \rightarrow R$ e $K : X \times X \rightarrow R$. Il successivo insieme di iperparametri ($\mathbf{x}_{next} \in X$) da valutare durante il processo di ottimizzazione viene determinato risolvendo l'equazione $\mathbf{x}_{next} = \operatorname{argmax}_{\mathbf{x}} a(\mathbf{x})$. La funzione $a : X \rightarrow R^+$ prende il nome di funzione di acquisizione e solitamente dipende sia dalle osservazioni analizzate in precedenza $\{\mathbf{x}^{(o)}, \mathbf{y}^{(o)}\}_{o=1}^O$ che dai parametri θ di GP. La formulazione è la seguente: $a(\mathbf{x}; \{\mathbf{x}^{(o)}, \mathbf{y}^{(o)}\}, \theta)$. Nell'impostazione del GP a priori, $a(\cdot)$ si basa sulle funzioni predittive di media e varianza del modello GP definite rispettivamente come $\mu(\mathbf{x}; \{\mathbf{x}^{(o)}, \mathbf{y}^{(o)}\}, \theta)$ e $\sigma^2(\mathbf{x}; \{\mathbf{x}^{(o)}, \mathbf{y}^{(o)}\}, \theta)$. In letteratura si possono trovare diverse definizioni di $a(\cdot)$ [Kushner, 1964, Mockus et al., 1978], ma l'*Upper Confidence Bound* (UCB) descritto da [Srinivas et al., 2009] ha dimostrato di essere il più efficiente per quanto riguarda il numero di valutazioni necessarie al fine di identificare l'ottimo globale di diverse funzioni multi-modali. La funzione di acquisizione è definita come $a_{UCB}(\mathbf{x}; \{\mathbf{x}^{(o)}, \mathbf{y}^{(o)}\}, \theta) = \mu(\mathbf{x}; \{\mathbf{x}^{(o)}, \mathbf{y}^{(o)}\}, \theta) - \kappa \sigma(\mathbf{x}; \{\mathbf{x}^{(o)}, \mathbf{y}^{(o)}\}, \theta)$, con κ a rappresentare un iperparametro che garantisce l'equilibrio tra sfruttamento dell'iperspazio degli iperparametri (ovvero favorire le parti che il modello identifica come promettenti) ed esplorazione dello stesso.

3.8 Valutazione della bontà di adattamento

Le metriche di valutazione della bontà di adattamento sono indissolubilmente legate ai processi di apprendimento automatico. Esistono metriche specifiche per ciascun problema, sia esso di classificazione, regressione, ranking, clustering, ecc. Tuttavia, esistono anche metriche trasversali che possono essere utilizzate nella valutazione dei risultati di problemi diversi.

All'interno di questa trattazione, l'attenzione sarà esclusivamente focalizzata sulle metriche utilizzate per analizzare i risultati ottenuti da modelli di apprendimento supervisionato.

3.8.1 Metriche di classificazione

Come anticipato al paragrafo 3.2.3, la classificazione consiste nel prevedere le etichette delle classi in base ai dati di input. Per ragioni legate alla semplificazione della trattazione, di seguito soltanto un breve cenno verrà fatto alle metriche utilizzate nei problemi di classificazione binaria. Tuttavia, tali metriche possono essere semplicemente ed efficacemente estese allo scenario multiclasse. A titolo esemplificativo, un classico esempio di classificazione binaria è il rilevamento dello spam, dove i dati di input possono includere il testo dell'e-mail ed i metadati (mittente, ora di invio, ecc.) e l'etichetta di output è "spam" o "non spam". Alcune volte vengono assegnati nomi generici alle due classi, come ad esempio: "positivo" e "negativo", oppure "classe 1" e "classe 0". Esistono molti modi per valutare le prestazioni dei modelli nel risolvere problemi di classificazione. Tuttavia, tra le metriche più diffuse troviamo:

- Accuratezza (*accuracy*), una semplice misura della frequenza con cui il classificatore esegue una previsione corretta. Rappresenta il rapporto tra il numero di previsioni corrette ed il numero totale di previsioni effettuate (il numero di osservazioni contenute nel set di test).

$$\text{Accuratezza} = \frac{\# \text{ predizioni corrette}}{\# \text{ numero totale di predizioni}} \quad (3.8.1)$$

- Matrice di confusione (*confusion matrix*), una metrica che cerca di sopperire ad una grave lacuna dell'accuratezza. Infatti, nonostante sia una metrica semplice ed intuitiva, l'accuratezza non fa alcuna distinzione tra le classi. Le risposte corrette per la classe 0 e per la classe 1 sono trattate allo stesso modo ma questo, talvolta, potrebbe non essere sufficiente. Infatti, si potrebbe voler esaminare il numero di classificazioni fallite per la classe 0 rispetto alla classe 1, poiché il costo dell'errore di classificazione potrebbe essere molto diverso per le due classi (è il caso delle diagnosi mediche, dei falsi positivi e dei falsi negativi). Una matrice di confusione (o tabella di confusione) mostra quindi una ripartizione più dettagliata delle classificazioni corrette ed errate per ciascuna classe. Le righe della matrice corrispondono alle etichette delle osservazioni sperimentali, mentre le colonne rappresentano le predizioni. Supponiamo che il dataset di test contenga 100 esempi di classe positiva e 200 esempi di classe negativa; la tabella di confusione potrebbe avere un aspetto simile a questo:

Tabella 3.5: Esempio di matrice di confusione.

	Predetto come positivo	Predetto come negativo
Etichettato come positivo	80	20
Etichettato come negativo	5	195

Osservando la matrice, si nota chiaramente che la classe positiva ha un'accuratezza inferiore ($80/(20 + 80) = 80\%$) rispetto alla classe negativa ($195/(5 + 195) = 97,5\%$). Questa

informazione si perderebbe valutando esclusivamente l'accuratezza complessiva, che in questo caso risulterebbe pari a $(80 + 195)/(100 + 200) = 91,7\%$.

- Accuratezza per classe (*per-class accuracy*), una variazione della semplice accuratezza che si calcola come la media delle accuratezze per ciascuna classe. La semplice accuratezza è definita come micro-media, mentre l'accuratezza per classe è definita come macro-media. Nell'esempio precedente, l'accuratezza media per classe risulterebbe $(80\% + 97,5\%)/2 = 88,75\%$. Si noti che in questo caso l'accuratezza media per classe risulta piuttosto diversa dalla semplice accuratezza. In generale, quando c'è un numero diverso di esempi per classe, l'accuratezza per classe sarà diversa dalla semplice accuratezza. Tale fattore riveste notevole importanza poiché, quando le classi sono sbilanciate (ovvero sono presenti molti più esempi di una classe rispetto all'altra), la semplice misura di accuratezza restituisce un'immagine molto distorta. La classe con più esempi, infatti, dominerà la statistica. In casi come questo, è necessario considerare l'accuratezza per classe, sia quella media che quella individuale. L'accuratezza per classe non è priva di limitazioni. Ad esempio, se ci sono pochi esempi di una classe, le statistiche del test per quella determinata classe avranno un'ampia varianza. Pertanto, la stima dell'accuratezza non sarà affidabile come quella relativa ad altre classi. In generale, il calcolo dell'accuratezza media di tutte le classi nasconde le misure di attendibilità delle singole classi.
- Precisione-Recupero (*Precision-Recall*), sebbene siano due metriche distinte, solitamente vengono utilizzate in maniera accoppiata. La precisione rappresenta l'accuratezza delle sole predizioni positive, mentre il recupero rappresenta il rapporto di istanze positive correttamente rilevate dal classificatore. Avendo cura di definire i termini Vero Positivo (*True Positive* - TP), Vero Negativo (*True Negative* - TN), Falso Positivo (*False Positive* - FP) e Falso-Negativo (*False Negative* - FN) per confrontare la classificazione di un'istanza con la corretta classificazione desiderata, è possibile definire:

$$\text{Precisione} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3.8.2)$$

$$\text{Recupero} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3.8.3)$$

- Punteggio F1 (*F1-score*), una metrica convenientemente utilizzata per fondere Precisione e Recupero in un unico punteggio. Il punteggio F1 rappresenta infatti la media armonica tra Precisione e Recupero, calcolata come il reciproco della media aritmetica dei reciproci:

$$\text{F1} = \frac{\text{TP}}{\text{TP} + \frac{\text{FN} + \text{FP}}{2}} \quad (3.8.4)$$

3.8.2 Metriche di regressione

Una modellazione predittiva in termini di regressione consiste nell'approssimare una funzione di mappatura (f) che, a partire da alcune variabili di input (X), restituisce una (o più - in

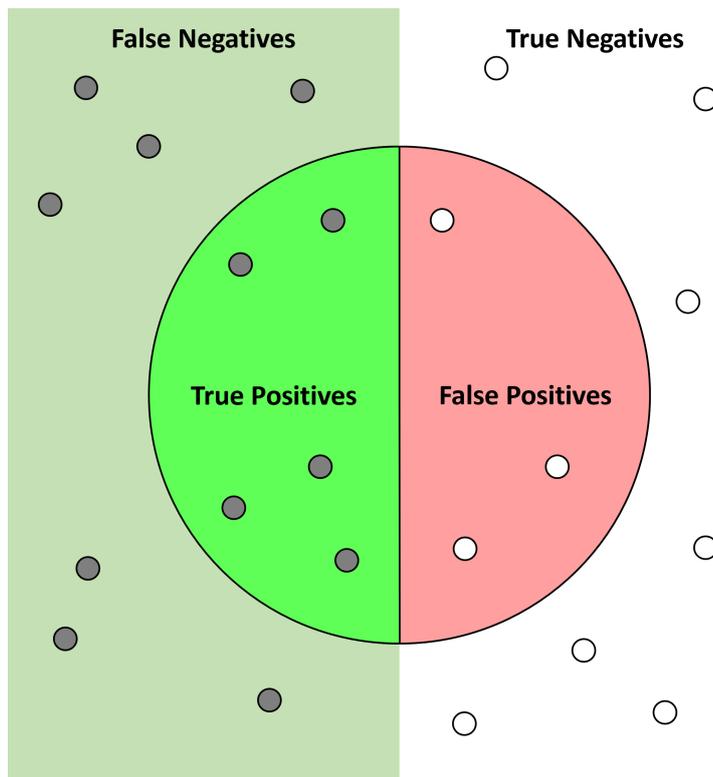


Figura 3.10: Schematizzazione grafica degli elementi valutati nella Precisione-recupero.

questo caso si parla di multi-regressione) variabile numerica continua di output (y). In questo caso, quindi, non è più possibile parlare di accuratezza. Non ha senso chiedere che un modello preveda esattamente un determinato valore (come ad esempio un'altezza, un importo in euro ecc.). Piuttosto, si è interessati nel comprendere quanto le previsioni si avvicinino ai valori attesi. Le capacità e le prestazioni di un modello di regressione devono dunque essere riportate in termini di errore nelle previsioni. Anche in questo caso, esistono molti modi per valutare le prestazioni dei modelli nel risolvere problemi di regressione. Avendo cura di definire con y_{T_i} e y_{P_i} rispettivamente l' i -esima osservazione target e l' i -esimo valore predetto, con n il numero complessivo di osservazioni, con μ e σ rispettivamente la media e la deviazione standard, tra le metriche più diffuse troviamo:

- Errore Quadratico Medio (*Mean Squared Error* - MSE), una metrica molto diffusa nei contesti di regressione poiché viene spesso utilizzata anche come funzione di perdita dagli algoritmi di addestramento in un inquadramento ai minimi quadrati. L'MSE si calcola come la media delle differenze al quadrato tra i valori target previsti e quelli attesi. L'elevamento a potenza sortisce l'effetto di dare più peso agli errori più grandi. In questo senso, maggiore è la differenza tra valore previsto e quello atteso, maggiore sarà l'errore

positivo al quadrato risultante.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{T_i} - y_{P_i})^2 \quad (3.8.5)$$

- Radice dell'Errore Quadratico Medio (*Root Mean Squared Error* - RMSE), una metrica molto comune e considerata un'eccellente misura dell'errore per predizioni numeriche. Si calcola come la radice quadrata della media del quadrato di tutti gli errori. Mantiene l'unità di misura della variabile in esame; pertanto, essendo dipendente dalla scala, è un'ottima misura delle performance di un modello e permette di confrontare gli errori in fase di predizione di diversi modelli (o configurazioni di uno stesso modello) ma soltanto per una specifica variabile e non tra variabili diverse.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{T_i} - y_{P_i})^2} \quad (3.8.6)$$

- Errore Assoluto Medio (*Mean Absolute Error* - MAE), una metrica molto comune poiché la sua unità di misura corrisponde a quella della variabile di target. Il MAE si calcola come la media dei valori assoluti degli errori commessi in fase predittiva. Pertanto, la differenza tra valore atteso e previsto può essere sia positiva che negativa; tuttavia, nel calcolo del MAE tale differenza diventerà necessariamente positiva. In questo caso, i diversi errori non vengono ponderati in maniera diversa, in quanto la metrica aumenta linearmente all'aumentare degli errori.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{T_i} - y_{P_i}| \quad (3.8.7)$$

- Errore Percentuale Assoluto Medio (*Mean Absolute Percentage Error* - MAPE), una metrica utile ed intuitiva che fornisce una misura di errore relativo utilizzando i valori assoluti affinché errori positivi e negativi non si annullino a vicenda. A titolo esemplificativo, un valore MAPE del 20% indica che la differenza percentuale assoluta media tra previsioni e valori reali è pari al 20%. In altre parole, le previsioni del modello si discostano in media del 20% dai valori reali. Chiaramente, un valore MAPE più basso è indicatore di previsioni più accurate (la condizione ottimale è rappresentata da un valore MAPE pari a 0%), mentre un valore MAPE più alto è indicatore di previsioni meno accurate.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_{T_i} - y_{P_i}}{y_{T_i}} \right| \cdot 100 \quad (3.8.8)$$

- Coefficiente di correlazione di Pearson (*Pearson Correlation Coefficient* - R), un coefficiente che misura la correlazione lineare tra due serie di dati. È definito come il rapporto tra la covarianza di due variabili ed il prodotto delle loro deviazioni standard. Essendo

sostanzialmente una misura normalizzata della covarianza, il suo risultato varia nel range $[-1;+1]$. Il segno "+" indica una correlazione positiva (se una variabile aumenta, aumenta anche l'altra), mentre il segno "-" indica una correlazione negativa (se una variabile aumenta, l'altra diminuisce). Il valore assoluto costituisce una misura della forza della correlazione. Gli estremi dell'intervallo rappresentano una correlazione perfetta, il che significa che è possibile determinare il valore esatto di una variabile a partire dal valore dell'altra. D'altra parte, un valore assoluto pari a 0 significa che non esiste alcuna relazione lineare tra le due variabili in esame: conoscere il valore di una variabile non aiuta in alcun modo a prevedere il valore dell'altra.

$$R = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{y_{T_i} - \mu_{y_{T_i}}}{\sigma_{y_{T_i}}} \right) \left(\frac{y_{P_i} - \mu_{y_{P_i}}}{\sigma_{y_{P_i}}} \right) \quad (3.8.9)$$

- Coefficiente di determinazione (*Coefficient of Determination* - R^2), una misura statistica che rappresenta la proporzione della varianza di una variabile dipendente attribuibile ad una variabile indipendente in un modello di regressione. Mentre la correlazione è indice della forza della relazione tra una variabile indipendente ed una dipendente, l' R^2 determina in che misura la varianza di una variabile è attribuibile alla varianza della seconda variabile.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{T_i} - y_{P_i})^2}{\sum_{i=1}^n (y_{T_i} - \mu_{y_{T_i}})^2} \quad (3.8.10)$$

Capitolo 4

Algoritmi di Machine Learning

” *If you invent a breakthrough in artificial intelligence, so machines can learn, that is worth 10 Microsofts.*

– Bill Gates

4.1 Regressione Lineare

La regressione lineare rappresenta una tecnica statistica utilizzata per comprendere la relazione esistente tra una variabile dipendente (talvolta definita risposta) ed una o più variabili indipendenti (talvolta definiti predittori). L'obiettivo principale della regressione lineare è modellare tale relazione attraverso un'equazione lineare che possa essere utilizzata per fare previsioni o comprendere la natura della connessione tra le variabili coinvolte. Indicando con y la variabile target che dovrà essere predetta dal modello e con (x_1, \dots, x_n) il set di variabili indipendenti, una regressione lineare ha lo scopo di identificare quei parametri $(\beta_0, \dots, \beta_n)$ in modo tale che $\beta_0 + \sum_{i=1}^n \beta_i x_i$ risulti un'affidabile approssimazione di y . Per ciascuna variabile x_i , un dataset contiene solitamente un numero m di osservazioni indicate con $x_{i,j}$. Pertanto, i parametri β_i vengono selezionati al fine di identificare il minimo della quantità denominata somma residua dei quadrati (*Residual Sum of Squares* - RSS) espressa come segue:

$$RSS = \sum_{j=1}^m \left(\beta_0 + \sum_{i=1}^n \beta_i x_{i,j} - y_j \right)^2 \quad (4.1.1)$$

Al fine di identificare una procedura che permettesse di identificare i valori ottimali dei coefficienti β_i , è stato formalmente introdotto [Hastie et al., 2009] il metodo di selezione graduale in avanti (*forward stepwise selection method*, [Bendel and Afifi, 1977]). Questo consiste nell'assumere β_0 pari al valore medio di y , mentre tutti gli altri β_i sono posti pari a 0. Successivamente, i ed i corrispondenti β_i vengono iterativamente selezionati in modo da produrre un modello capace di ridurre al minimo la quantità RSS. Una volta scelto, ciascun valore di β_i resta invariato per la rimanente durata della procedura. Conseguentemente, ogni modello successivo si differenzia da quello precedente per un singolo parametro β . Infine, saranno prodotti $n + 1$ modelli differenti, ciascuno risultante in un valore iterativamente più basso della quantità RSS.

4.1.1 Regressione Ridge

La regressione Ridge (*Ridge Regression* - RR) rappresenta una particolare versione di un modello di regressione lineare giacché, ancor prima di regolare i valori dei parametri β , una procedura di normalizzazione viene applicata ai valori delle variabili predittive. Tale procedura consiste nel determinare la media e la deviazione standard di ciascuna variabile contenuta nel set di dati. Successivamente, per ciascuna di queste, occorre sottrarre la media al valore della singola osservazione e dividere il tutto per il valore di deviazione standard. In questo modo, tutte le variabili saranno riportate alla stessa scala ed avranno una deviazione standard pari a 1. Un'ulteriore peculiarità della regressione Ridge consiste nell'applicare una penalità ai coefficienti β_i di modo che siano preferibili quelli con valori ridotti. Lo scopo ultimo è quello di ridurre i valori dei coefficienti associati alle variabili predittive meno importanti, mantenendo alti i β_i associati alle variabili predittive più importanti. Di conseguenza, la funzione di costo da ottimizzare durante una regressione Ridge (CF_{RR}) è leggermente diversa dalla RSS analizzata in precedenza, in virtù della penalizzazione introdotta. La sua formulazione matematica risulta:

$$CF_{RR} = RSS + \alpha \sum_{j=1}^m \beta_j^2 \quad (4.1.2)$$

dove α rappresenta il parametro positivo di penalizzazione che determina un compromesso tra la minimizzazione dell'errore ed il valore contenuto dei coefficienti di regressione. Chiaramente, la scelta di diversi valori di α genera la messa a punto di modelli differenti.

4.1.2 Regressione Lasso

La regressione Lasso (*Lasso Regression* - LR) si differenzia dalla regressione Ridge appena descritta in virtù di un'ulteriore modifica della funzione di costo (CF_{LR}) [Friedman et al., 2010]. La sua formulazione matematica risulta:

$$CF_{LR} = RSS + \alpha \sum_{j=1}^m |\beta_j| \quad (4.1.3)$$

In questo caso, infatti, il termine di penalizzazione α viene applicato alla sommatoria del valore assoluto dei coefficienti di penalizzazione β_j . Il vantaggio apportato da questa leggera modifica consiste nella capacità di poter rendere nulli alcuni coefficienti di regressione qualora il valore di α sia molto elevato. L'annullamento di alcuni coefficienti di regressione conduce a modelli di regressione lineare più contenuti, in virtù del minor numero di β . Se ne deduce che α sia l'unico iperparametro di entrambe le regressioni Ridge e Lasso; pertanto, durante la messa a punto di un modello di apprendimento automatico, sarà necessario ottimizzarlo attraverso le tecniche descritte al paragrafo 3.7.

4.2 Macchine a Vettori di Supporto

Le macchine a vettori di supporto (*Support Vector Machines* - SVM) rappresentano un potente strumento per risolvere sia problemi di classificazione che di regressione nell'ambito dell'apprendimento automatico supervisionato. Queste vengono comunemente impiegate nei diversi settori

dell'ingegneria [Yang et al., 2022, Babagoli and Rezaei, 2022] e, qualora siano implementate per problemi di regressione, prendono il nome specifico di SVR - *Support Vector Regressor*. Tali algoritmi sono particolarmente indicati nell'approcciare dataset limitati e/o problemi ad elevata dimensionalità [Cortes and Vapnik, 1995].

Sia assegnato un set di dati (x_i, y_i) con $i = 1, 2, \dots, n$ in cui x_i corrisponde al vettore delle variabili nello spazio degli input mentre y_i rappresenta il corrispondente valore target richiesto in output. L'obiettivo principale di un SVR è quello di identificare un iperpiano di regressione $f(x)$ che minimizzi la distanza tra le osservazioni target ed il piano stesso. La funzione di confine decisionale si determina come segue:

$$f(x) = \omega^T \phi(x) + b \quad (4.2.1)$$

in cui il vettore ω e l'intercetta b controllano la posizione dell'iperpiano di confine decisionale.

Il principio di funzionamento di un SVR viene graficamente mostrato in Figura 4.1. Si introduce una funzione di perdita L_ϵ determinata in funzione del margine ϵ . Il valore di L_ϵ è diverso da 0 se e solo se le osservazioni si trovano al di fuori dell'intervallo $f(x) \pm \epsilon$. Il modello viene quindi ottimizzato in maniera tale da massimizzare l'ampiezza dell'intervallo, minimizzando così il valore della funzione di perdita.

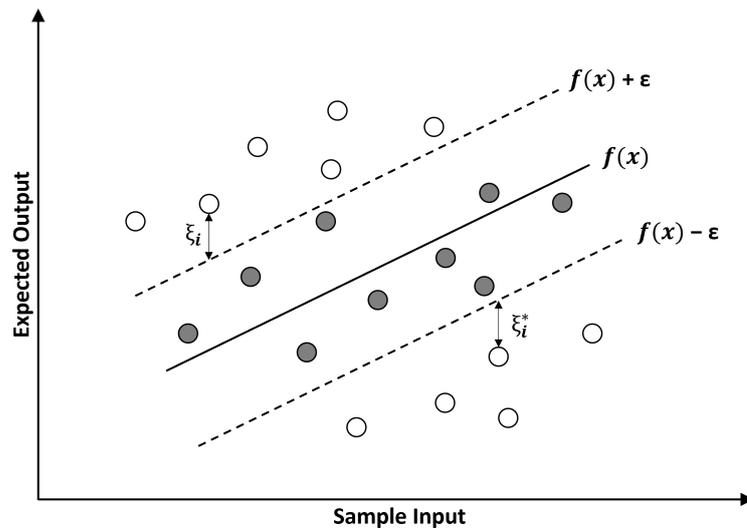


Figura 4.1: Principio di funzionamento di un SVR.

4.2.1 Kernel Trick

Se le osservazioni sperimentali non sono linearmente separabili nello spazio originale, è possibile applicare delle trasformazioni in modo tale da mappare i dati dello spazio originale in uno spazio delle caratteristiche di dimensioni maggiori. Tale pratica è particolarmente consolidata e prende il nome di *kernel trick*. In linea generale, la funzione di perdita si definisce come:

$$L_\epsilon(f(x_i) - y_i) = \begin{cases} 0 & \text{se } |f(x_i) - y_i| \leq \epsilon \\ |f(x_i) - y_i| - \epsilon & \text{se } |f(x_i) - y_i| > \epsilon \end{cases} \quad (4.2.2)$$

In base al principio di minimizzazione del rischio strutturale [Vapnik, 1991], il coefficiente di penalizzazione C e le variabili di rilassamento (*slack variables*) ξ_i e ξ_i^* vengono introdotte per trasformare il problema di regressione in un problema di ottimizzazione vincolata di funzione:

$$\arg \min_{\omega, b} \frac{\|\omega\|^2}{2} + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (4.2.3)$$

di modo che, per $i = 1, 2, \dots, n$, vengano rispettate le seguenti condizioni:

$$\begin{cases} (\omega^T x_i + b) - y_i \leq \epsilon + \xi_i \\ y_i - (\omega^T x_i + b) \leq \epsilon + \xi_i^* \\ \xi_i \geq 0 \\ \xi_i^* \geq 0 \end{cases} \quad (4.2.4)$$

Il coefficiente di penalizzazione C viene utilizzato per controllare la tolleranza degli errori del modello. Tanto minore sarà il valore di C , tanto maggiori saranno i valori ammissibili di ξ_i e ξ_i^* e, di conseguenza, tanto maggiore sarà la tolleranza per gli errori. Introducendo il metodo dei moltiplicatori di Lagrange [Nguyen et al., 2012], la funzione obiettivo si trasforma in un problema duale:

$$\min \left\{ \frac{1}{2} \sum_{i=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) - \sum_{i=1}^N \alpha_i (y_i - \epsilon) + \sum_{i=1}^N \alpha_i^* (y_i + \epsilon) \right\} \quad (4.2.5)$$

con:

$$\begin{cases} \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \quad (4.2.6)$$

Il principio KKT (o condizione Karush-Kuhn-Tucker [Gordon and Tibshirani, 2012]) viene quindi impiegato per risolvere il problema di regressione in esame. La funzione di regressione del modello predittivo SVR diventa quindi:

$$f(x) = \sum_{i=1}^K (\alpha_i^* - \alpha_i) K(x_i, x_j) + b \quad (4.2.7)$$

con α_i^* e α_i che rappresentano i fattori di Lagrange e $K(x_i, x_j)$ a rappresentare la corrispondente funzione kernel. Le funzioni kernel più comunemente impiegate sono: lineare (Equazione 4.2.8), polinomiale (Equazione 4.2.9), gaussiana (Equazione 4.2.10) e sigmoideale (Equazione 4.2.11):

$$K(x_i, x_j) = x_i^T x_j \quad (4.2.8)$$

$$K(x_i, x_j) = (-\gamma x_i^T x_j + 1)^d \quad (4.2.9)$$

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (4.2.10)$$

$$K(x_i, x_j) = \tanh(-\gamma x_i^T x_j + 1)^d \quad (4.2.11)$$

con γ e d che rappresentano i parametri delle funzioni kernel da ottimizzare in maniera opportuna e \tanh a rappresentare la funzione tangente iperbolica.

4.3 Reti Neurali Artificiali

4.3.1 Analogia biologica

Una rete neurale artificiale (*Artificial Neural Network* - ANN) rappresenta un modello di apprendimento automatico evolutosi dall'idea di simulare il funzionamento del cervello umano. L'elaborazione delle informazioni da parte del cervello umano avviene attraverso componenti biologiche che operano in parallelo per produrre funzionalità adeguate, come il pensiero o l'apprendimento. La cellula fondamentale del sistema nervoso centrale è il neurone, il cui ruolo consiste nel trasmettere impulsi (stimoli elettrici originati da reazioni fisico-chimiche) in determinate condizioni di funzionamento. Questo elemento biologico può essere suddiviso in tre componenti principali: i dendriti, il corpo cellulare (noto anche come "soma") e l'assone (Figura 4.2).

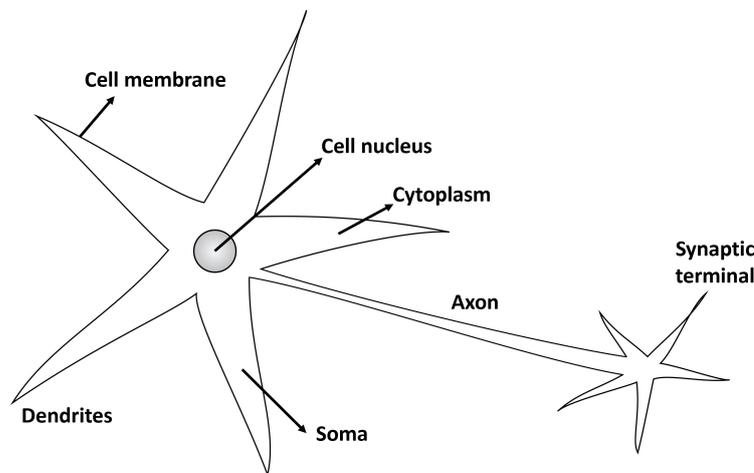


Figura 4.2: Schematizzazione di un neurone biologico.

I dendriti sono composti da diversi prolungamenti sottili che formano l'albero dendritico. Lo scopo fondamentale dei dendriti è quello di acquisire, in modo continuo, stimoli da altri neuroni (connettori) o dall'ambiente esterno, come nel caso dei neuroni a contatto con l'ambiente (chiamati anche sensoriali). Il corpo cellulare, oltre a contenere i principali organuli citoplasmatici (nucleo, mitocondri, lisosoma ecc.), è responsabile dell'elaborazione di tutte le informazioni provenienti dai dendriti, al fine di produrre un potenziale di attivazione che

indica se il neurone può attivare o meno un impulso elettrico lungo il suo assone. L'assone è composto da un unico prolungamento che ha il compito di trasmettere gli impulsi elettrici verso altri neuroni di collegamento o verso neuroni direttamente connessi al tessuto muscolare. La terminazione dell'assone è composta anche da rami chiamati terminali sinaptici. Le sinapsi sono le connessioni che permettono il trasferimento degli impulsi elettrici assonici da un particolare neurone ai dendriti di altri neuroni, come illustrato in Figura 4.3.

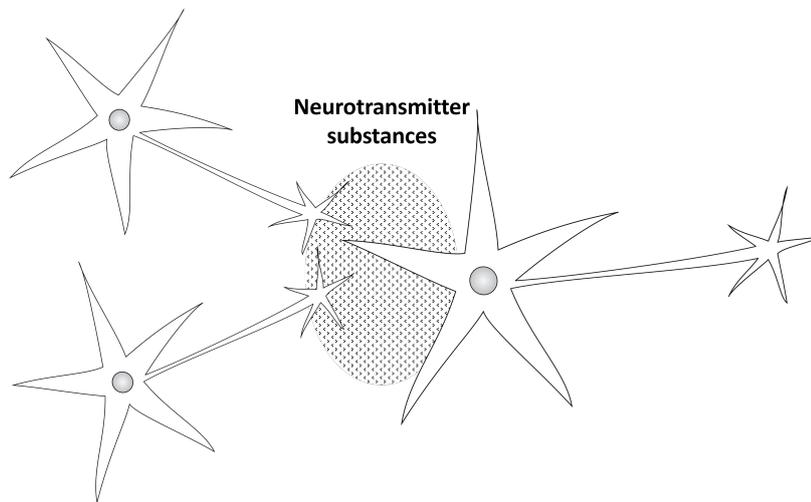


Figura 4.3: Rappresentazione delle connessioni sinaptiche tra neuroni.

È importante sottolineare che non vi sia un vero contatto fisico tra i neuroni che formano la giunzione sinaptica. Gli elementi neurotrasmettitoriali rilasciati sulla giunzione hanno infatti il compito di pesare la trasmissione da un neurone all'altro. La funzionalità di un neurone dipende dai suoi pesi sinaptici, che sono anch'essi dinamici e dipendenti dalla chimica cerebrale [Hodgkin and Huxley, 1952]. In sintesi, sebbene le attività legate al neurone biologico possano sembrare inizialmente molto semplici, i suoi componenti, quando funzionano complessivamente, sono responsabili di tutte le elaborazioni effettuate e gestite dal cervello umano. Si stima che la rete neurale biologica sia composta da circa 100 miliardi (10^{11}) di neuroni. Ognuno di essi è interconnesso attraverso connessioni sinaptiche (rese possibili da più di cinquanta sostanze neurotrasmettitorie) a una media di 6000 neuroni, per un totale di 600 trilioni di sinapsi [Shepherd, 2003]. Il potenziale d'azione della membrana neurale ha valori negativi quando è a riposo (polarizzata), il che significa che si verifica una maggiore concentrazione di ioni negativi all'interno della membrana rispetto all'esterno. Quando la cellula nervosa viene stimolata (depolarizzata) con un impulso superiore alla sua soglia di attivazione (-55 mV), causato dalla variazione delle concentrazioni interne di ioni sodio (Na^+) e potassio (K^+), innesca un impulso elettrico che si propagerà lungo tutto il suo assone con un'ampiezza massima di 35 mV [Schwartz et al., 1991]. Le fasi relative alle variazioni del potenziale d'azione all'interno di un neurone durante la sua stimolazione sono mostrate in Figura 4.4. È importante sottolineare che il valore massimo del potenziale d'azione è fissato a 35 mV per tutti i neuroni sottoposti a

stimolazione: tuttavia, la durata del segnale nel tempo è variabile. Tale condizione si verifica indipendentemente dalla categoria di neurone. Al termine del processo di stimolazione, la membrana si ripolarizza ed il potenziale d'azione ritorna alla tensione di riposo (-70 mV), come illustrato in Figura 4.4.

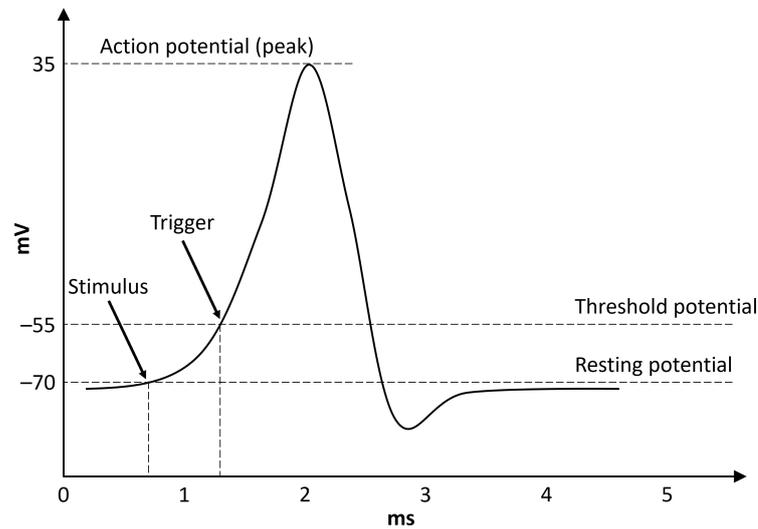


Figura 4.4: Stadi del potenziale d'azione.

4.3.2 Neurone artificiale

Le strutture delle reti neurali artificiali sono state sviluppate a partire da modelli noti di sistemi nervosi biologici e dello stesso cervello umano. Le componenti computazionali o unità di elaborazione, chiamate neuroni artificiali, sono modelli semplificati dei neuroni biologici, ispirati dall'analisi di come la membrana cellulare di un neurone generi e propaghi gli impulsi elettrici [Hodgkin and Huxley, 1952]. I neuroni artificiali utilizzati nelle reti neurali artificiali sono solitamente non-lineari, forniscono uscite continue e svolgono funzioni semplici come raccogliere i segnali disponibili in input, assemblarli in base alle loro caratteristiche funzionali e produrre una risposta in base ad una funzione di attivazione. Il modello di neurone più semplice ma comprensivo delle caratteristiche principali di una rete neurale biologica (parallelismo ed elevata connettività) è stato proposto nel 1943 [McCulloch and Pitts, 1943] ed è tuttora il più utilizzato nelle diverse architetture di reti neurali artificiali. In questo modello, ogni neurone di una rete può essere schematizzato come in Figura 4.5.

I diversi segnali di input provenienti dall'ambiente esterno sono rappresentati dall'insieme $\{x_1, x_2, x_3, \dots, x_n\}$, in analogia agli impulsi elettrici esterni raccolti dai dendriti nel neurone biologico. I pesi delle giunzioni sinaptiche della rete sono implementati all'interno del neurone artificiale mediante l'insieme dei pesi sinaptici $\{w_1, w_2, w_3, \dots, w_n\}$. Pertanto, l'importanza di ciascuna variabile in ingresso al neurone x_i si ottiene moltiplicando il valore in ingresso per il corrispondente peso w_i , ponderando così tutte le informazioni esterne che arrivano al

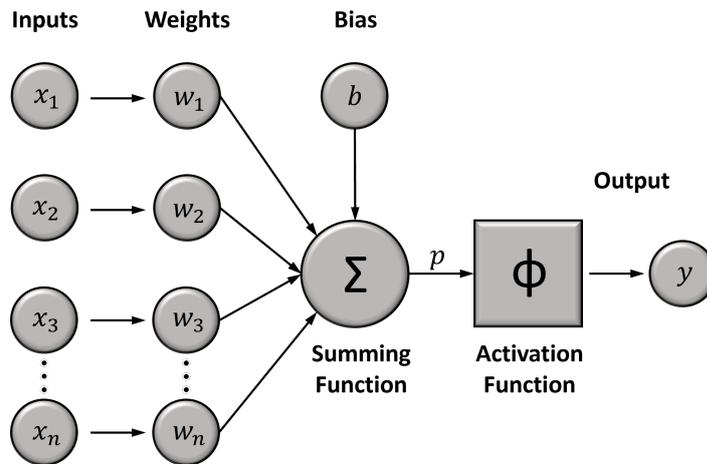


Figura 4.5: Schematizzazione di un neurone artificiale.

neurone. Prendendo in considerazione la Figura 4.5, si può osservare come il neurone artificiale sia composto da sette elementi fondamentali, ossia:

1. Segnali di Input ($x_1, x_2, x_3, \dots, x_n$), rappresentano i segnali provenienti dall'ambiente esterno ovvero i valori assunti dalle variabili di una particolare banca dati. Come descritto al paragrafo 3.3.5, i segnali di input solitamente subiscono una normalizzazione al fine di migliorare l'efficienza computazionale dell'algoritmo di apprendimento;
2. Pesi sinaptici ($w_1, w_2, w_3, \dots, w_n$), rappresentano i valori utilizzati per ponderare ciascuno dei segnali di input: consentono di quantificarne la rilevanza rispetto alle funzionalità del neurone;
3. Sommatore (Σ), rappresenta l'elemento che raccoglie tutti i segnali ponderati di input al fine di produrre un unico segnale;
4. Bias o soglia di attivazione (b), rappresenta una variabile utilizzata per specificare la soglia che il segnale prodotto dal sommatore deve superare affinché venga generato un innesco (*trigger*) per la produzione di un determinato output;
5. Potenziale di attivazione (p), rappresenta il risultato della differenza tra il segnale prodotto dal sommatore e la soglia di attivazione. Se tale valore risulta positivo, ovvero $p \geq b$, allora il neurone produce un segnale eccitatorio. In caso contrario, ovvero $p < b$, il segnale prodotto sarà inibitorio;
6. Funzione di attivazione (ϕ), il cui obiettivo è limitare l'ampiezza del segnale in uscita dal neurone entro un intervallo ragionevole di valori;
7. Segnale di Output (y), rappresenta il valore finale prodotto dal neurone a partire da un particolare insieme di segnali di input: può essere utilizzato anche come input per altri neuroni sequenzialmente interconnessi.

Pertanto, è possibile riepilogare il funzionamento del neurone artificiale proposto da McCulloch e Pitts attraverso i seguenti passaggi operativi:

- fornire al neurone una serie di valori (a rappresentare le variabili in input);
- moltiplicare ciascun ingresso del neurone per il corrispondente peso sinaptico;
- calcolare il potenziale di attivazione come il prodotto della somma ponderata dei singoli segnali di ingresso e sottrarre la soglia di attivazione;
- applicare un'opportuna funzione di attivazione per limitare l'ampiezza del segnale in uscita dal neurone;
- elaborare l'output.

L'output y prodotto dal neurone artificiale può dunque essere elaborato mediante le Equazioni 4.3.1 e 4.3.2:

$$p = \sum_{i=1}^n w_i \cdot x_i - b \quad (4.3.1)$$

$$y = \phi(p) \quad (4.3.2)$$

4.3.3 Gestione delle non linearità

La funzione di attivazione rappresenta un elemento fondamentale per la gestione delle non linearità: senza di essa, infatti, il neurone diventerebbe soltanto un semplice sommatore lineare. In generale, in funzione dei corrispondenti domini di esistenza, le funzioni di attivazione possono essere classificate in due gruppi principali:

1. Funzioni di attivazione parzialmente differenziabili, ovvero funzioni con punti in cui risultano nulle le derivate del primo ordine. Le tre principali funzioni appartenenti a questa categoria sono:
 - Funzione a gradino (*Threshold function* oppure *Heaviside function*), assume un valore unitario positivo quando il potenziale di attivazione del neurone è maggiore o uguale a zero. In caso contrario, il valore assunto dalla funzione a gradino sarà nullo

$$\phi(p) = \begin{cases} 1 & \text{se } p \geq 0 \\ 0 & \text{se } p < 0 \end{cases} \quad (4.3.3)$$

- Funzione segnale o funzione bipolare a gradino (*Symmetric hard limiter*), assume un valore unitario positivo quando il potenziale di attivazione del neurone è maggiore di zero, un valore nullo quando il potenziale è nullo ed un valore unitario negativo quando il potenziale di attivazione è minore di zero

$$\phi(p) = \begin{cases} 1 & \text{se } p > 0 \\ 0 & \text{se } p = 0 \\ -1 & \text{se } p < 0 \end{cases} \quad (4.3.4)$$

- Funzione rampa simmetrica (*Symmetric ramp function*), assume un valore pari al potenziale di attivazione quando definito all'interno dell'intervallo $[-a, a]$ ed assume valori pari agli estremi dell'intervallo all'esterno di questo

$$\phi(p) = \begin{cases} a & \text{se } p > a \\ p & \text{se } -a \leq p \leq a \\ -a & \text{se } p < -a \end{cases} \quad (4.3.5)$$

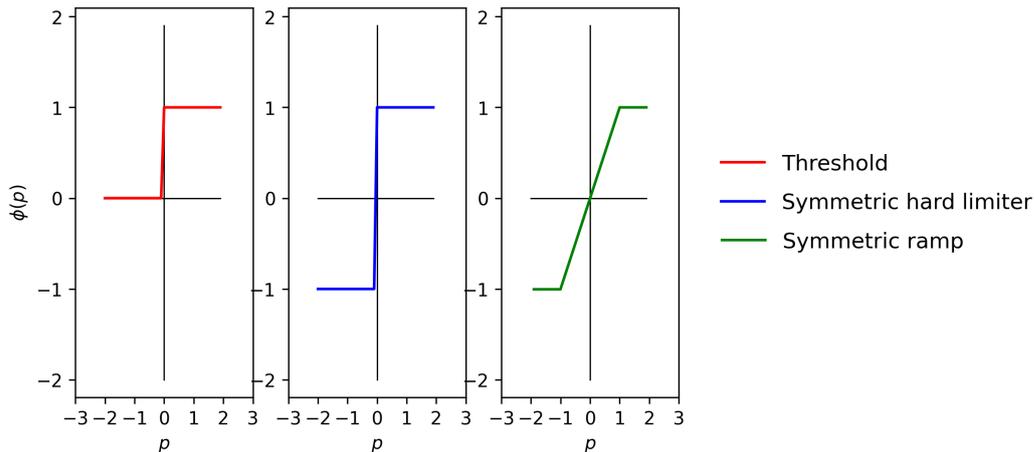


Figura 4.6: Funzioni di attivazione parzialmente differenziabili ($a = 1$).

2. Funzioni di attivazione totalmente differenziabili, ovvero funzioni per le quali le derivate del primo ordine esistono e sono definite per tutti i punti del dominio di esistenza della funzione. Le quattro principali funzioni appartenenti a questa categoria (ed utilizzate all'interno delle reti neurali artificiali) sono:

- Funzione identità (*Identity function*), assume un valore pari al potenziale di attivazione per ciascun punto del dominio

$$\phi(p) = p \quad (4.3.6)$$

- Funzione logistica (*Logistic function*), assume valori reali compresi tra 0 e 1

$$\phi(p) = \frac{1}{1 + e^{-\beta \cdot p}} \quad (4.3.7)$$

β è una costante reale associata alla pendenza della funzione nel suo punto di flesso. Se ne può concludere che la rappresentazione grafica della funzione di attivazione logistica è molto simile a quella della funzione a gradino, soprattutto per valori di β molto alti e tendenti ad infinito. Tuttavia, a differenza della funzione a gradino, la funzione logistica è completamente differenziabile all'interno del dominio di esistenza.

- Funzione tangente iperbolica (*Hyperbolic tangent function*), assume valori reali compresi tra -1 e 1

$$\phi(p) = \frac{1 - e^{-\beta \cdot p}}{1 + e^{-\beta \cdot p}} \quad (4.3.8)$$

anche in questo caso, β è una costante reale associata alla pendenza della funzione nel suo punto di flesso. Per valori di β molto alti e tendenti ad infinito, la funzione tangente iperbolica tende alla funzione bipolare a gradino. Sia la funzione logistica che quella tangente iperbolica appartengono ad una famiglia di funzioni che prende il nome di sigmoidali.

- Funzione Gaussiana (*Gaussian function*), assume valori uguali per potenziali di attivazione posti alla stessa distanza dal centro (media). La curva è simmetrica rispetto al centro e la funzione Gaussiana è data da

$$\phi(p) = e^{-\frac{(p-c)^2}{2\sigma^2}} \quad (4.3.9)$$

con c e σ che definiscono rispettivamente il centro della funzione Gaussiana e la deviazione standard associata, ovvero la dispersione della curva rispetto al suo centro.

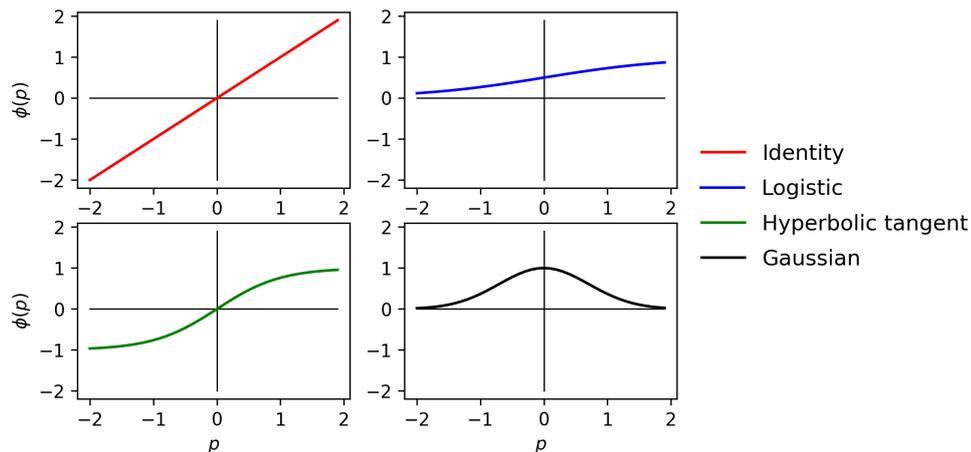


Figura 4.7: Funzioni di attivazione totalmente differenziabili ($\beta = 1$).

4.3.4 Architettura di una rete neurale

L'architettura di una rete neurale artificiale definisce il modo in cui i vari neuroni sono disposti e messi in relazione l'uno con l'altro. Tali disposizioni sono essenzialmente articolate attraverso la direzione delle connessioni sinaptiche e definiscono la cosiddetta topologia della rete. La topologia è infatti definita attraverso la composizione strutturale che una rete neurale è in grado di assumere: tuttavia, al variare del numero di neuroni, è possibile avere topologie differenti pur mantenendo costante l'architettura della rete. In linea generale, è possibile suddividere una

rete neurale artificiale in tre principali gruppi funzionali che prendono il nome di strati. Questi sono rispettivamente noti come:

1. Strato di ingresso (*Input layer*), responsabile della ricezione delle informazioni, dei segnali, delle variabili provenienti dall'ambiente esterno. Solitamente tali dati vengono normalizzati all'interno dei valori limite prodotti dalle funzioni di attivazione in virtù di una migliore precisione numerica nelle operazioni matematiche eseguite dalla rete;
2. Strato nascosto (*Hidden layer*), luogo in cui avviene l'effettiva elaborazione computazionale in funzione di una specifica funzione di attivazione in grado di modulare l'ampiezza del segnale in uscita dai neuroni appartenenti a questo strato. È opportuno aggiungere più strati nascosti man mano che aumenta la complessità del problema da analizzare. Tuttavia per risolvere la maggior parte dei problemi di regressione è sufficiente implementare un singolo strato nascosto composto da un adeguato numero di neuroni [Demuth et al., 2014]. Tale architettura prende il nome di rete neurale superficiale (*Shallow Neural Network - SNN*). In caso contrario, sarà opportuno progettare un modello di rete neurale composto da più strati nascosti realizzando un'architettura che prende il nome di rete neurale profonda (*Deep Neural Network - DNN*);
3. Strato di uscita (*Output layer*), responsabile della produzione e della rappresentazione degli output finali della rete, risultato dell'elaborazione effettuata dai neuroni degli strati precedenti.

4.3.5 Principali architetture neurali

Le principali architetture di rete neurale artificiale, considerando la disposizione dei neuroni, nonché il modo in cui sono interconnessi e come sono composti i suoi strati, possono essere sintetizzate in:

- Reti *feedforward* a singolo strato, in cui, escluso lo strato di input, è presente un unico strato neurale rappresentato dallo strato di output. In Figura 4.8 è possibile osservare un esempio di rete *feedforward* a singolo strato composto da n ingressi e m uscite. Le informazioni fluiscono sempre in un'unica direzione (unidirezionale), partendo dallo strato di ingresso ed arrivando a quello di uscita. Per questo particolare tipo di architettura, il numero di outputs corrisponde esattamente al numero di neuroni della rete. La tipologia principale di rete *feedforward* a singolo strato prende il nome di Percettrone, il cui algoritmo di apprendimento utilizzato in fase di addestramento si basa sulla regola di Hebb [Hebb, 1949].
- Reti *feedforward* multistrato, in cui sono presenti uno o più strati neurali nascosti (Figura 4.9). Tali architetture sono solitamente impiegate nella risoluzione di diversi problemi, dall'approssimazione di funzioni alla classificazione di modelli, dal controllo di processi all'ottimizzazione e così via. In Figura 4.9 è possibile osservare un esempio di rete *feedforward* multistrato composto da n ingressi, m uscite e due strati neurali nascosti formati rispettivamente da n_1 ed n_2 neuroni. La tipologia principale di rete *feedforward* multistrato prende il nome di Percettrone multistrato (*MultiLayer Perceptron - MLP*), il cui

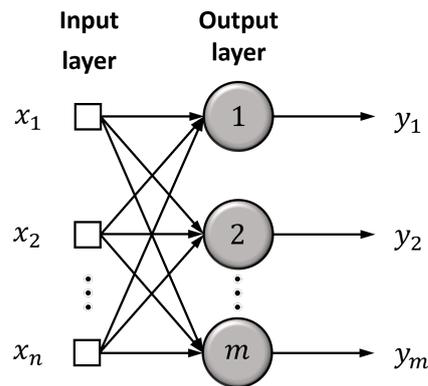


Figura 4.8: Esempio di una rete feedforward a singolo strato.

algoritmo di apprendimento utilizzato in fase di addestramento si basa sulla regola delta generalizzata [Rumelhart et al., 1986]. Solitamente, il numero di neuroni che compongono il primo strato nascosto differisce dal numero di segnali che compongono lo strato di input della rete. Infatti, il numero di strati nascosti ed il rispettivo numero di neuroni dipendono dalla natura e dalla complessità del fenomeno in esame, oltre che dalla quantità e dalla qualità dei dati disponibili al riguardo. Tuttavia, analogamente a quanto accade per le reti *feedforward* a singolo strato, la quantità di segnali di output coincide ancora con il numero di neuroni presenti nel rispettivo strato.

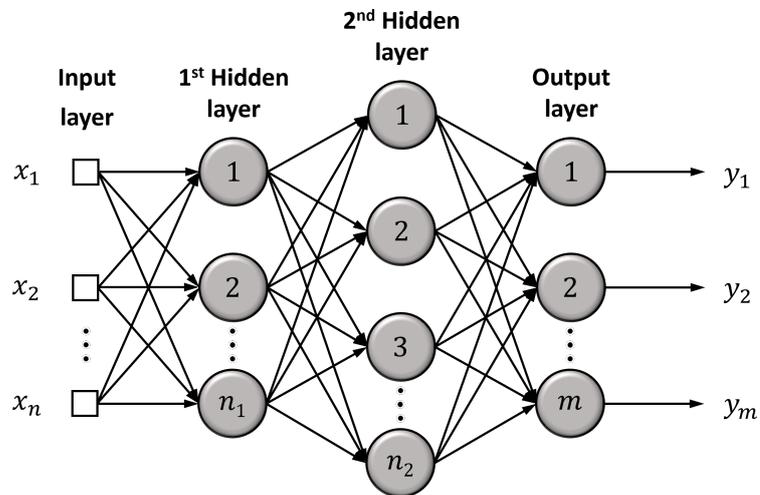


Figura 4.9: Esempio di una rete feedforward multistrato.

- Reti ricorrenti, in cui i segnali in uscita dai neuroni sono utilizzati come *feedback* in ingresso di altri neuroni. Tale caratteristica permette alle reti ricorrenti di elaborare le informazioni in maniera dinamica e di essere impiegate per l'analisi di sistemi variabili

nel tempo (previsione di serie temporali). La tipologia principale di rete ricorrente prende il nome di Percettrone con *feedback* tra neuroni di strati distinti, il cui algoritmo di apprendimento utilizzato in fase di addestramento si basa ancora sulla regola delta generalizzata [Rumelhart et al., 1986]. In Figura 4.10 è possibile osservare un esempio di Percettrone con *feedback*, in cui uno dei segnali in uscita viene reintrodotta in uno strato intermedio. In questo modo, attraverso un processo di retroazione, i segnali di uscita tengono conto anche dei valori di output precedentemente prodotti.

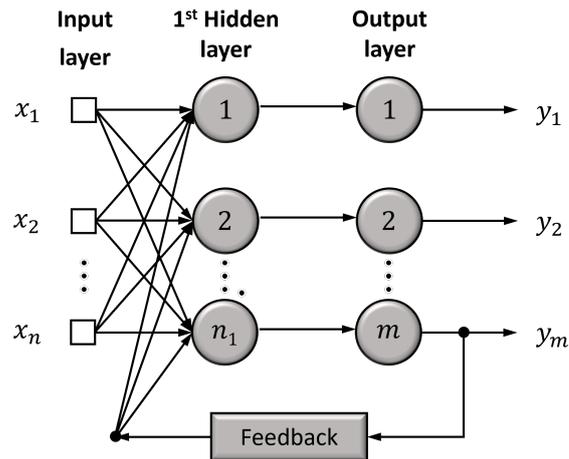


Figura 4.10: Esempio di una rete ricorrente.

4.3.6 Percettrone multistrato

Le reti MLP sono particolarmente rinomate per il loro ampio spettro di applicazioni nelle diverse aree della conoscenza. Sono infatti considerate una delle architetture più versatili in termini di applicabilità [Silva et al., 2016]. Analizzando la Figura 4.9, è possibile osservare come ogni segnale di input della rete venga propagato strato dopo strato verso quello di output. In questo caso, gli output dei neuroni del primo strato nascosto costituiranno gli input dei neuroni del secondo strato nascosto. Analogamente, gli output dei neuroni del secondo strato nascosto costituiranno gli input dei neuroni dello strato di output. Pertanto, la propagazione dei segnali di un MLP è unidirezionale, dallo strato di input verso quello di output. Come accennato al paragrafo 4.3.2, i pesi e le soglie di ciascun neurone di una rete MLP necessitano di essere ottimizzati mediante un processo di addestramento supervisionato. L'algoritmo di apprendimento utilizzato durante l'addestramento di un MLP si chiama retropropagazione (*backpropagation*). Esso consiste in due principali passaggi: uno in avanti ed uno all'indietro (Figura 4.11). Il primo prende il nome di propagazione in avanti (*forward pass*) e consiste nel fornire i segnali $\{x_1, x_2, \dots, x_n\}$ in input alla rete per propagarli strato dopo strato fino alla produzione degli output corrispondenti. Questa fase mira unicamente ad ottenere le risposte dalla rete, tenendo conto dei valori attuali dei pesi sinaptici e delle soglie dei singoli neuroni. Tali valori resteranno inalterati durante l'esecuzione di questa fase. Successivamente, poiché si tratta

di un processo di apprendimento supervisionato, le risposte prodotte in output dalla rete vengono confrontate con le risposte auspiccate (le osservazioni disponibili). Le differenze tra le risposte auspiccate e quelle ottenute dai neuroni in uscita saranno utilizzate al fine di regolare i singoli pesi e le soglie di ciascun neurone. Pertanto, proprio a causa di questi errori in fase predittiva, risulta necessario eseguire il passaggio all'indietro (*backward pass*). A differenza di quanto accade nel passaggio in avanti, in questo caso vengono applicate delle correzioni per regolare i valori dei pesi sinaptici e delle soglie di tutti i neuroni della rete. In sintesi, l'applicazione successiva dei passaggi in avanti e all'indietro consente di correggere automaticamente i pesi sinaptici e le soglie dei neuroni iterazione dopo iterazione, ottenendo una graduale riduzione dell'errore complessivo commesso dalla rete rispetto alle risposte auspiccate.

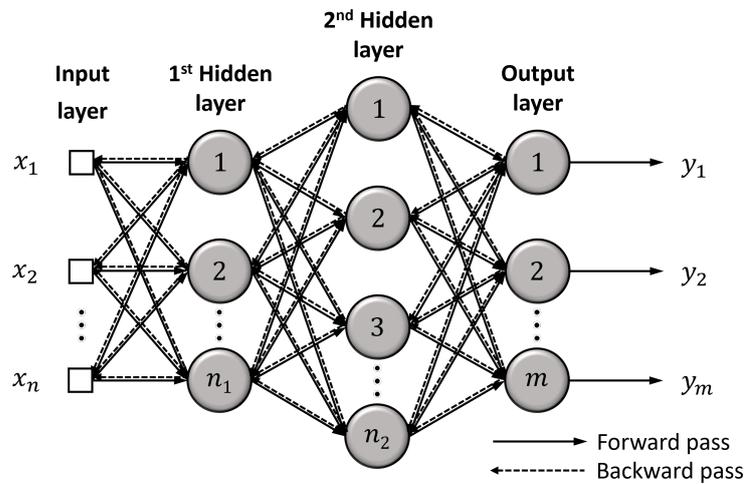


Figura 4.11: Schematizzazione dei passaggi dell'algoritmo di retropropagazione.

4.3.7 Algoritmo di retropropagazione

Per una comprensione migliore del principio di funzionamento dell'algoritmo di retropropagazione, è necessario definire alcune variabili e parametri ausiliari che verranno utilizzati a tal proposito. Sulla base della topologia MLP illustrata in Figura 4.11, è possibile identificare l'insieme delle variabili che governano l'algoritmo di retropropagazione. Ogni neurone j appartenente ad uno degli strati S della topologia rappresentata in Figura 4.12 può essere configurato attraverso una funzione di attivazione $f(\cdot)$ continua e differenziabile in tutto il suo dominio (come quelle descritte al punto 2 del paragrafo 4.3.3).

Siano:

- $W_{ji}^{(S)}$ la matrice dei pesi i cui elementi rappresentano i valori dei pesi sinaptici delle connessioni tra il j -esimo neurone dello strato S e l' i -esimo neurone dello strato $S - 1$. Nello specifico, in riferimento alla topologia mostrata in Figura 4.12 è possibile identificare:

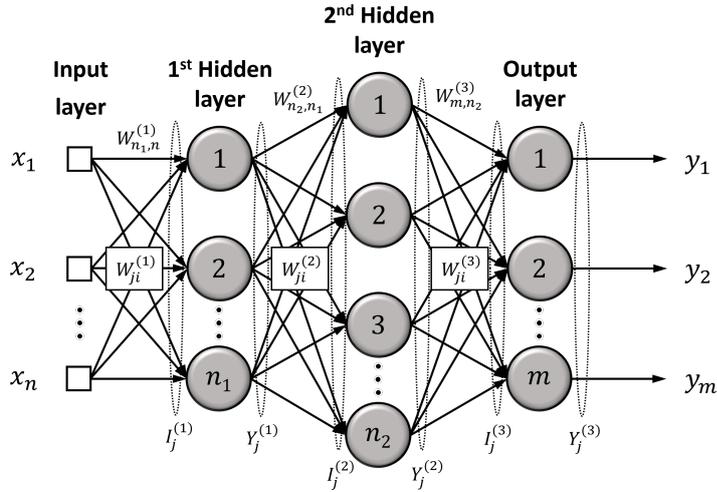


Figura 4.12: Notazione utile alla derivazione dell'algoritmo di retropropagazione.

- $W_{ji}^{(3)}$, matrice dei pesi delle connessioni tra il j -esimo neurone dello strato di output all' i -esimo neurone dello strato nascosto 2;
 - $W_{ji}^{(2)}$, matrice dei pesi delle connessioni tra il j -esimo neurone dello strato nascosto 2 all' i -esimo neurone dello strato nascosto 1;
 - $W_{ji}^{(1)}$, matrice dei pesi delle connessioni tra il j -esimo neurone dello strato nascosto 1 all' i -esimo segnale dello strato di input.
- $I_j^{(S)}$ i vettori i cui elementi denotano gli ingressi pesati relativi al j -esimo neurone dello strato S definiti mediante le equazioni:

$$I_j^{(1)} = \sum_{i=0}^n W_{ji}^{(1)} \cdot x_i = W_{1,0}^{(1)} \cdot x_0 + W_{1,1}^{(1)} \cdot x_1 + \dots + W_{1,n}^{(1)} \cdot x_n \quad (4.3.10)$$

$$I_j^{(2)} = \sum_{i=0}^{n_1} W_{ji}^{(2)} \cdot Y_i^{(1)} = W_{1,0}^{(2)} \cdot Y_0^{(1)} + W_{1,1}^{(2)} \cdot Y_1^{(1)} + \dots + W_{1,n_1}^{(2)} \cdot Y_{n_1}^{(1)} \quad (4.3.11)$$

$$I_j^{(3)} = \sum_{i=0}^{n_2} W_{ji}^{(3)} \cdot Y_i^{(2)} = W_{1,0}^{(3)} \cdot Y_0^{(2)} + W_{1,1}^{(3)} \cdot Y_1^{(2)} + \dots + W_{1,n_2}^{(3)} \cdot Y_{n_2}^{(2)} \quad (4.3.12)$$

- $Y_j^{(S)}$ i vettori i cui elementi denotano l'uscita del j -esimo neurone in relazione allo strato S definiti mediante le equazioni:

$$Y_j^{(1)} = f(I_j^{(1)}) \quad (4.3.13)$$

$$Y_j^{(2)} = f(I_j^{(2)}) \quad (4.3.14)$$

$$Y_j^{(3)} = f(I_j^{(3)}) \quad (4.3.15)$$

Il passo successivo nella derivazione dell'algoritmo di retropropagazione consiste nel definire una funzione che rappresenti l'errore di approssimazione, il cui scopo è misurare la deviazione delle risposte prodotte dai neuroni di uscita della rete rispetto ai valori auspicati. Considerando il k -esimo campione di osservazioni di addestramento e la topologia illustrata in Figura 4.12, la funzione più utilizzata per misurare le performance associate ai risultati prodotti dai neuroni di output è l'errore quadratico:

$$E(k) = \frac{1}{2} \sum_{j=1}^m \left(a_j(k) - Y_j^{(3)}(k) \right)^2 \quad (4.3.16)$$

avendo cura di indicare con $Y_j^{(3)}(k)$ il valore prodotto dal j -esimo neurone di output della rete per il k -esimo campione di osservazioni di addestramento, e con $a_j(k)$ il rispettivo valore auspicato. Di conseguenza, assumendo un training set composto da p campioni, la misura delle performance globali dell'algoritmo di retropropagazione può essere calcolata attraverso l'errore quadratico medio:

$$E_M = \frac{1}{p} \sum_{k=1}^p E(k) \quad (4.3.17)$$

La regolazione dei pesi sinaptici si differenzia a seconda che si consideri lo strato di output oppure uno strato intermedio. Pertanto, per migliorare la comprensibilità della trattazione, seguiranno due diversi paragrafi in cui tali descrizioni saranno rispettivamente affrontate.

4.3.8 Regolazione dei pesi sinaptici dello strato di output

L'obiettivo principale del processo di addestramento dello strato di output consiste nel regolare la matrice dei pesi $W_{ji}^{(3)}$ in maniera tale da minimizzare l'errore tra le uscite prodotte dalla rete rispetto a quelle auspicate. Considerando l'equazione 4.3.16, la definizione di gradiente e la regola della catena per la derivazione delle funzioni composte, è possibile scrivere:

$$\nabla E^{(3)} = \frac{\partial E}{\partial W_{ji}^{(3)}} = \frac{\partial E}{\partial Y_j^{(3)}} \cdot \frac{\partial Y_j^{(3)}}{\partial I_j^{(3)}} \cdot \frac{\partial I_j^{(3)}}{\partial W_{ji}^{(3)}} \quad (4.3.18)$$

Sulla base delle precedenti definizioni, è possibile scrivere:

$$\frac{\partial I_j^{(3)}}{\partial W_{ji}^{(3)}} = Y_i^{(2)} \quad (4.3.19)$$

$$\frac{\partial Y_j^{(3)}}{\partial I_j^{(3)}} = f'(I_j^{(3)}) \quad (4.3.20)$$

$$\frac{\partial E}{\partial Y_j^{(3)}} = -(a_j - Y_j^{(3)}) \quad (4.3.21)$$

dove f' rappresenta la derivata prima della funzione di attivazione implementata. Combinando tali espressioni, è possibile ottenere:

$$\frac{\partial E}{\partial W_{ji}^{(3)}} = -\left(a_j - Y_j^{(3)}\right) \cdot f'\left(I_j^{(3)}\right) \cdot Y_i^{(2)} \quad (4.3.22)$$

Pertanto, la regolazione della matrice dei pesi $W_{ji}^{(3)}$ deve avvenire nella direzione opposta al gradiente in maniera tale da minimizzare l'errore. Risulterà quindi:

$$\Delta W_{ji}^{(3)} = -\eta \cdot \frac{\partial E}{\partial W_{ji}^{(3)}} \quad (4.3.23)$$

$$\Delta W_{ji}^{(3)} = \eta \cdot \delta_j^{(3)} \cdot Y_i^{(2)} \quad (4.3.24)$$

avendo definito $\delta_j^{(3)}$ come il gradiente locale relativo al j -esimo neurone dello strato di output, ovvero:

$$\delta_j^{(3)} = \left(a_j - Y_j^{(3)}\right) \cdot f'\left(I_j^{(3)}\right) \quad (4.3.25)$$

Inoltre, è possibile convertire l'equazione 4.3.24 nella seguente procedura iterativa:

$$W_{ji}^{(3)}(t+1) = W_{ji}^{(3)}(t) + \eta \cdot \delta_j^{(3)} \cdot Y_i^{(2)} \quad (4.3.26)$$

dove η rappresenta il tasso di apprendimento dell'algoritmo di retropropagazione. In questo senso, l'equazione 4.3.26 permette di regolare i pesi delle connessioni sinaptiche dello strato di output tenendo conto delle differenze osservate tra le risposte prodotte dalla rete ed i rispettivi valori auspicati.

4.3.9 Regolazione dei pesi sinaptici degli strati intermedi

A differenza dei neuroni appartenenti allo strato di output di un MLP, i neuroni degli strati intermedi non hanno accesso diretto ai valori auspicati per le loro uscite. In questo caso, la regolazione dei loro pesi sinaptici avviene attraverso una stima degli errori di uscita prodotti dai neuroni dello strato successivo (già regolato, per definizione dell'algoritmo di retropropagazione). A titolo esemplificativo, osservando sempre la topologia proposta in Figura 4.12, la correzione dei pesi dei neuroni del secondo strato intermedio avverrà solo dopo aver concluso la regolazione dei neuroni dello strato di output. Proprio in questo aspetto risiede l'essenza dell'algoritmo di retropropagazione, poiché in un primo momento i pesi sinaptici dei neuroni dello strato di output vengono regolati confrontando le risposte prodotte dalla rete con i rispettivi valori auspicati. In un secondo momento, tale errore viene retropropagato ai neuroni degli strati precedenti, ponderandolo con i valori dei pesi sinaptici precedentemente regolati in tutti gli strati successivi. Di conseguenza, la risposta auspicata di un neurone di uno strato nascosto deve essere determinata rispetto ai neuroni dello strato immediatamente successivo che sono direttamente collegati ad esso e che sono già stati regolati nel precedente passaggio. Pertanto:

$$\nabla E^{(2)} = \frac{\partial E}{\partial W_{ji}^{(2)}} = \frac{\partial E}{\partial Y_j^{(2)}} \cdot \frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} \cdot \frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} \quad (4.3.27)$$

Sulla base delle precedenti definizioni, è possibile scrivere:

$$\frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} = Y_i^{(1)} \quad (4.3.28)$$

$$\frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} = f'(I_j^{(2)}) \quad (4.3.29)$$

$$\frac{\partial E}{\partial Y_j^{(2)}} = \sum_{k=1}^m \frac{\partial E}{\partial I_k^{(3)}} \cdot \frac{\partial I_k^{(3)}}{\partial Y_j^{(2)}} \quad (4.3.30)$$

L'equazione 4.3.30 può essere riscritta come:

$$\frac{\partial E}{\partial Y_j^{(2)}} = \sum_{k=1}^m \frac{\partial E}{\partial I_k^{(3)}} \cdot \frac{\partial \left(\sum_{k=1}^m W_{kj}^{(3)} \cdot Y_j^{(2)} \right)}{\partial Y_j^{(2)}} \quad (4.3.31)$$

Semplificando, è possibile scrivere:

$$\frac{\partial E}{\partial Y_j^{(2)}} = \sum_{k=1}^m \frac{\partial E}{\partial I_k^{(3)}} \cdot W_{kj}^{(3)} \quad (4.3.32)$$

$W_{kj}^{(3)}$ si riferisce al peso sinaptico di tutti i neuroni dello strato di uscita che sono interconnessi ad un dato neurone j del secondo strato intermedio. Per sottolineare ancora una volta l'essenza dell'algoritmo di retropropagazione, si può notare come tutti i pesi $W_{kj}^{(3)}$ siano già stati regolati nel passaggio precedente sulla base degli errori realmente commessi e, pertanto, saranno utilizzati per regolare i pesi del secondo strato intermedio. Il primo membro del prodotto nell'equazione 4.3.32 può essere ottenuto mediante moltiplicazione delle equazioni 4.3.20 e 4.3.21. Il risultato sarà il valore attuale di $\delta_j^{(3)}$ ottenuto dall'equazione 4.3.25. Pertanto:

$$\frac{\partial E}{\partial Y_j^{(2)}} = - \sum_{k=1}^m \delta_k^{(3)} \cdot W_{kj}^{(3)} \quad (4.3.33)$$

Di conseguenza:

$$\frac{\partial E}{\partial W_{ji}^{(2)}} = - \sum_{k=1}^m \delta_k^{(3)} \cdot W_{kj}^{(3)} \cdot f'(I_j^{(2)}) \cdot Y_i^{(1)} \quad (4.3.34)$$

Pertanto, la regolazione della matrice dei pesi $W_{ji}^{(2)}$ deve avvenire nella direzione opposta al gradiente in maniera tale da minimizzare l'errore. Risulterà quindi:

$$\Delta W_{ji}^{(2)} = -\eta \cdot \frac{\partial E}{\partial W_{ji}^{(2)}} \quad (4.3.35)$$

$$\Delta W_{ji}^{(2)} = \eta \cdot \delta_j^{(2)} \cdot Y_i^{(1)} \quad (4.3.36)$$

avendo definito $\delta_j^{(2)}$ come il gradiente locale relativo al j -esimo neurone del secondo strato intermedio, ovvero:

$$\delta_j^{(2)} = - \left(\sum_{k=1}^m \delta_k^{(3)} \cdot W_{kj}^{(3)} \right) \cdot f' \left(I_j^{(2)} \right) \quad (4.3.37)$$

Convertendo l'equazione 4.3.36 in una procedura iterativa:

$$W_{ji}^{(2)}(t+1) = W_{ji}^{(2)}(t) + \eta \cdot \delta_j^{(2)} \cdot Y_i^{(1)} \quad (4.3.38)$$

In questo senso, l'equazione 4.3.38 permette di regolare i pesi delle connessioni sinaptiche del secondo strato nascosto propagando all'indietro l'errore originato dai neuroni dello strato nascosto.

Analogamente, è possibile regolare i pesi delle connessioni sinaptiche del primo strato nascosto propagando all'indietro l'errore originato dai neuroni del secondo strato nascosto, mediante le equazioni:

$$\delta_j^{(1)} = - \left(\sum_{k=1}^{n_2} \delta_k^{(2)} \cdot W_{kj}^{(2)} \right) \cdot f' \left(I_j^{(1)} \right) \quad (4.3.39)$$

$$W_{ji}^{(1)}(t+1) = W_{ji}^{(1)}(t) + \eta \cdot \delta_j^{(1)} \cdot x_i \quad (4.3.40)$$

Chiaramente, le procedure per la regolazione delle matrici dei pesi sinaptici descritte nelle equazioni precedenti possono essere generalizzate per qualsiasi topologia di rete MLP, indipendentemente dal numero di strati intermedi.

4.3.10 Algoritmo Levenberg-Marquardt

Come illustrato in precedenza, l'algoritmo di retropropagazione regola i valori delle matrici dei pesi nella direzione opposta al gradiente della funzione di perdita. Tuttavia, all'atto pratico, questo algoritmo tende a convergere molto lentamente, richiedendo quindi un elevato onere computazionale. Per evitare questo inconveniente e ridurre sia i tempi di convergenza che gli sforzi computazionali, sono state incorporate diverse tecniche di ottimizzazione all'algoritmo standard di retropropagazione. Tra le tante tecniche utilizzate a questo scopo, si distingue l'algoritmo di Levenberg-Marquardt (LM) [Hagan and Menhaj, 1994]. L'algoritmo LM è un metodo del gradiente di secondo ordine, basato sul metodo dei minimi quadrati per modelli non lineari, e può essere incorporato nell'algoritmo di retropropagazione per migliorare l'efficienza del processo di addestramento. Combinando le equazioni 4.3.16 e 4.3.17 si ottiene:

$$\begin{aligned}
 V &= \frac{1}{2p} \sum_{k=1}^p \sum_{j=1}^m \left(a_j(k) - Y_j^{(3)}(k) \right)^2 \\
 &= \frac{1}{2p} \sum_{k=1}^p \left(\mathbf{a}(k) - \mathbf{Y}^{(3)}(k) \right) \left(\mathbf{a}(k) - \mathbf{Y}^{(3)}(k) \right) \\
 &= \frac{1}{2p} \sum_{k=1}^p \mathbf{E}^T(k) \cdot \mathbf{E}(k)
 \end{aligned} \tag{4.3.41}$$

dove il termine $\{\mathbf{E}(k) = \mathbf{a}(k) - \mathbf{Y}^{(3)}(k)\}$ rappresenta il vettore dell'errore commesso considerando il k -esimo campione di osservazioni di addestramento. Per una particolare osservazione k , l'errore si ottiene mediante:

$$V = \frac{1}{2} \cdot \mathbf{E}^T(k) \cdot \mathbf{E}(k) \tag{4.3.42}$$

Mentre l'algoritmo di retropropagazione standard consiste nell'utilizzare il gradiente della funzione errore quadratico al fine di minimizzarla, l'algoritmo LM consiste in un'approssimazione del metodo di Newton [Battiti, 1992, Foresee and Hagan, 1997]. Pertanto, è possibile ottenere la minimizzazione di una funzione $V(z)$ rispetto ad un vettore parametrico z attraverso la seguente procedura iterativa:

$$\Delta z = -(\nabla^2 V(z))^{-1} \cdot \nabla V(z) \tag{4.3.43}$$

con $\nabla^2 V(z)$ e $\nabla V(z)$ a rappresentare rispettivamente la matrice Hessiana (matrice delle derivate del secondo ordine) e la matrice Jacobiana (matrice delle derivate del primo ordine) di $V(z)$. Supponendo che $V(z)$ sia una funzione che riproduca t funzioni quadratiche come quelle descritte nell'equazione 4.3.41, per un assegnato vettore parametrico z composto da q elementi, si ottiene la seguente espressione:

$$V(z) = \sum_{i=1}^t e_i^2(z) \tag{4.3.44}$$

A partire dall'equazione precedente, si può dimostrare come:

$$\nabla V(z) = \mathbf{J}^T(z) \cdot \mathbf{e}(z) \tag{4.3.45}$$

$$\nabla^2 V(z) = \mathbf{J}^T(z) \cdot \mathbf{J}(z) + \mu \cdot \mathbf{I} \tag{4.3.46}$$

essendo \mathbf{I} la matrice identità e μ un parametro utile a regolare il tasso di convergenza dell'algoritmo Levenberg-Marquardt. $\mathbf{J}(z)$ rappresenta la matrice Jacobiana, definita attraverso:

$$J(z) = \begin{bmatrix} \frac{\partial e_1(z)}{\partial z_1} & \frac{\partial e_1(z)}{\partial z_2} & \cdots & \frac{\partial e_1(z)}{\partial z_q} \\ \frac{\partial e_2(z)}{\partial z_1} & \frac{\partial e_2(z)}{\partial z_2} & \cdots & \frac{\partial e_2(z)}{\partial z_q} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_N(z)}{\partial z_1} & \frac{\partial e_N(z)}{\partial z_2} & \cdots & \frac{\partial e_N(z)}{\partial z_q} \end{bmatrix} \quad (4.3.47)$$

Combinando le equazioni 4.3.45 e 4.3.46 con l'equazione 4.3.43, è possibile ottenere l'equazione iterativa che governa il metodo LM:

$$\Delta z = \left(\mathbf{J}^T(z) \cdot \mathbf{J}(z) + \mu \cdot \mathbf{I} \right)^{-1} \cdot \mathbf{J}^T(z) \cdot \mathbf{e}(z) \quad (4.3.48)$$

Pertanto, la peculiarità di questo algoritmo risiede nel calcolo della matrice Jacobiana. Per il processo di addestramento delle reti MLP, come illustrato nella Figura 4.12, questa matrice Jacobiana (equazione 4.3.47) viene rielaborata rispetto alle matrici sinaptiche della rete:

$$\mathbf{J}(\mathbf{W}) = \left[\mathbf{J}(\mathbf{W}^{(1)}) \quad \mathbf{J}(\mathbf{W}^{(2)}) \quad \mathbf{J}(\mathbf{W}^{(3)}) \right] \quad (4.3.49)$$

essendo $\mathbf{J}(\mathbf{W}) \in R^{(p) \times ((n+1) \cdot n_1 + (n_1+1) \cdot n_2 + (n_2+1) \cdot m)}$.

Nel caso in esame, \mathbf{W} è composta da:

$$\begin{aligned} \mathbf{W} &= [\mathbf{W}^{(1)} \quad \mathbf{W}^{(2)} \quad \mathbf{W}^{(2)}] \\ &= \begin{bmatrix} W_{1,0}^{(1)} \cdots W_{1,n}^{(1)} & W_{2,0}^{(1)} \cdots W_{2,n}^{(1)} & \cdots & W_{n_1,0}^{(1)} \cdots W_{n_1,n}^{(1)} \\ W_{1,0}^{(2)} \cdots W_{1,n_1}^{(2)} & W_{2,0}^{(2)} \cdots W_{2,n_1}^{(2)} & \cdots & W_{n_2,0}^{(2)} \cdots W_{n_2,n_1}^{(2)} \\ W_{1,0}^{(3)} \cdots W_{1,n_2}^{(3)} & W_{2,0}^{(3)} \cdots W_{2,n_2}^{(3)} & \cdots & W_{m,0}^{(3)} \cdots W_{m,n_2}^{(3)} \end{bmatrix}^T \end{aligned} \quad (4.3.50)$$

dove $\mathbf{W} \in R^{((n+1) \cdot n_1 + (n_1+1) \cdot n_2 + (n_2+1) \cdot m)}$.

Di conseguenza, le matrici $\mathbf{J}(\mathbf{W}^{(1)})$, $\mathbf{J}(\mathbf{W}^{(2)})$ e $\mathbf{J}(\mathbf{W}^{(3)})$ saranno così definite:

$$\mathbf{J}(\mathbf{W}^{(1)}) = \begin{bmatrix} \frac{\partial \mathbf{E}(1)}{\partial W_{1,1}^{(1)}} \cdots \frac{\partial \mathbf{E}(1)}{\partial W_{1,n}^{(1)}} & \frac{\partial \mathbf{E}(1)}{\partial W_{2,1}^{(1)}} \cdots \frac{\partial \mathbf{E}(1)}{\partial W_{2,n}^{(1)}} & \cdots & \frac{\partial \mathbf{E}(1)}{\partial W_{n_1,1}^{(1)}} \cdots \frac{\partial \mathbf{E}(1)}{\partial W_{n_1,n}^{(1)}} \\ \frac{\partial \mathbf{E}(2)}{\partial W_{1,1}^{(1)}} \cdots \frac{\partial \mathbf{E}(2)}{\partial W_{1,n}^{(1)}} & \frac{\partial \mathbf{E}(2)}{\partial W_{2,1}^{(1)}} \cdots \frac{\partial \mathbf{E}(2)}{\partial W_{2,n}^{(1)}} & \cdots & \frac{\partial \mathbf{E}(2)}{\partial W_{n_1,1}^{(1)}} \cdots \frac{\partial \mathbf{E}(2)}{\partial W_{n_1,n}^{(1)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{E}(p)}{\partial W_{1,1}^{(1)}} \cdots \frac{\partial \mathbf{E}(p)}{\partial W_{1,n}^{(1)}} & \frac{\partial \mathbf{E}(p)}{\partial W_{2,1}^{(1)}} \cdots \frac{\partial \mathbf{E}(p)}{\partial W_{2,n}^{(1)}} & \cdots & \frac{\partial \mathbf{E}(p)}{\partial W_{n_1,1}^{(1)}} \cdots \frac{\partial \mathbf{E}(p)}{\partial W_{n_1,n}^{(1)}} \end{bmatrix} \quad (4.3.51)$$

$$\mathbf{J}(\mathbf{W}^{(2)}) = \begin{bmatrix} \frac{\partial \mathbf{E}(1)}{\partial W_{1,1}^{(2)}} \cdots \frac{\partial \mathbf{E}(1)}{\partial W_{1,n_1}^{(2)}} & \frac{\partial \mathbf{E}(1)}{\partial W_{2,1}^{(2)}} \cdots \frac{\partial \mathbf{E}(1)}{\partial W_{2,n_1}^{(2)}} & \cdots & \frac{\partial \mathbf{E}(1)}{\partial W_{n_2,1}^{(2)}} \cdots \frac{\partial \mathbf{E}(1)}{\partial W_{n_2,n_1}^{(2)}} \\ \frac{\partial \mathbf{E}(2)}{\partial W_{1,1}^{(2)}} \cdots \frac{\partial \mathbf{E}(2)}{\partial W_{1,n_1}^{(2)}} & \frac{\partial \mathbf{E}(2)}{\partial W_{2,1}^{(2)}} \cdots \frac{\partial \mathbf{E}(2)}{\partial W_{2,n_1}^{(2)}} & \cdots & \frac{\partial \mathbf{E}(2)}{\partial W_{n_2,1}^{(2)}} \cdots \frac{\partial \mathbf{E}(2)}{\partial W_{n_2,n_1}^{(2)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{E}(p)}{\partial W_{1,1}^{(2)}} \cdots \frac{\partial \mathbf{E}(p)}{\partial W_{1,n_1}^{(2)}} & \frac{\partial \mathbf{E}(p)}{\partial W_{2,1}^{(2)}} \cdots \frac{\partial \mathbf{E}(p)}{\partial W_{2,n_1}^{(2)}} & \cdots & \frac{\partial \mathbf{E}(p)}{\partial W_{n_2,1}^{(2)}} \cdots \frac{\partial \mathbf{E}(p)}{\partial W_{n_2,n_1}^{(2)}} \end{bmatrix} \quad (4.3.52)$$

$$\mathbf{J}(\mathbf{W}^{(3)}) = \begin{bmatrix} \frac{\partial \mathbf{E}(1)}{\partial W_{1,1}^{(3)}} \cdots \frac{\partial \mathbf{E}(1)}{\partial W_{1,n_2}^{(3)}} & \frac{\partial \mathbf{E}(1)}{\partial W_{2,1}^{(3)}} \cdots \frac{\partial \mathbf{E}(1)}{\partial W_{2,n_2}^{(3)}} & \cdots & \frac{\partial \mathbf{E}(1)}{\partial W_{m,1}^{(3)}} \cdots \frac{\partial \mathbf{E}(1)}{\partial W_{m,n_2}^{(3)}} \\ \frac{\partial \mathbf{E}(2)}{\partial W_{1,1}^{(3)}} \cdots \frac{\partial \mathbf{E}(2)}{\partial W_{1,n_2}^{(3)}} & \frac{\partial \mathbf{E}(2)}{\partial W_{2,1}^{(3)}} \cdots \frac{\partial \mathbf{E}(2)}{\partial W_{2,n_2}^{(3)}} & \cdots & \frac{\partial \mathbf{E}(2)}{\partial W_{m,1}^{(3)}} \cdots \frac{\partial \mathbf{E}(2)}{\partial W_{m,n_2}^{(3)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{E}(p)}{\partial W_{1,1}^{(3)}} \cdots \frac{\partial \mathbf{E}(p)}{\partial W_{1,n_2}^{(3)}} & \frac{\partial \mathbf{E}(p)}{\partial W_{2,1}^{(3)}} \cdots \frac{\partial \mathbf{E}(p)}{\partial W_{2,n_2}^{(3)}} & \cdots & \frac{\partial \mathbf{E}(p)}{\partial W_{m,1}^{(3)}} \cdots \frac{\partial \mathbf{E}(p)}{\partial W_{m,n_2}^{(3)}} \end{bmatrix} \quad (4.3.53)$$

dove $\mathbf{J}(\mathbf{W}^{(1)}) \in R^{(p) \times (n \cdot n_1)}$, $\mathbf{J}(\mathbf{W}^{(2)}) \in R^{(p) \times (n_1 \cdot n_2)}$ e $\mathbf{J}(\mathbf{W}^{(3)}) \in R^{(p) \times (n_2 \cdot m)}$.

A partire dall'equazione 4.3.48, l'espressione iterativa del metodo Levenberg-Marquardt che regola la matrice dei pesi della rete MLP si ridefinisce attraverso:

$$\Delta \mathbf{W} = \left[\mathbf{J}^T(\mathbf{W}) \cdot \mathbf{J}(\mathbf{W}) + \mu \mathbf{I} \right]^{-1} \cdot \mathbf{J}^T(\mathbf{W}) \cdot \mathbf{E} \quad (4.3.54)$$

avendo cura di definire con $\mathbf{E} = [\mathbf{E}(1)\mathbf{E}(2) \dots \mathbf{E}(p)]^T$ il vettore degli errori calcolati rispetto ai campioni p .

Infine, gli elementi delle matrici $\mathbf{J}(\mathbf{W}^{(1)})$, $\mathbf{J}(\mathbf{W}^{(2)})$ e $\mathbf{J}(\mathbf{W}^{(3)})$ saranno ottenuti in sequenza ripetendo i passaggi in avanti e all'indietro dell'algoritmo di retropropagazione precedentemente descritto. Attraverso le modifiche effettuate all'algoritmo standard, il metodo Levenberg-Marquardt ha dimostrato di velocizzare il processo di addestramento classico di un MLP ottenendo risultati in un tempo tra le 10 e le 100 volte inferiore rispetto all'algoritmo di retropropagazione convenzionale [Hagan and Menhaj, 1994].

4.4 Alberi Decisionali

Gli alberi decisionali (*Decision Trees* - DTs) rappresentano concettualmente e matematicamente un approccio diverso all'apprendimento automatico. Gli approcci finora descritti trattano dati strettamente numerici che possono aumentare e/o diminuire monotonamente. Le equazioni che definiscono questi approcci non sono in grado di elaborare dati che non siano numerici, ad esempio categorici o di tipo stringa. Tuttavia, la teoria degli alberi decisionali non si basa sul fatto che i dati siano numerici. Infatti, mentre gli altri approcci partono dalla formulazione di equazioni sulle proprietà dei dati, gli alberi decisionali iniziano con il disegno di una struttura ad albero in cui ad ogni nodo c'è una decisione da prendere. In sostanza, gli alberi decisionali sono

strutture euristiche che vengono costruite tramite una sequenza di scelte o confronti effettuati in un determinato ordine.

Gli alberi decisionali rappresentano uno strumento di regressione/classificazione potente e flessibile. In particolare:

- possono gestire variabili sia numeriche che categoriche in modo semplice e naturale;
- la classificazione/regressione finale ha una struttura semplice che può essere facilmente registrata ed è in grado di classificare/prevedere nuovi dati in modo efficiente;
- selezionano automaticamente le variabili in modo graduale, riducendo la complessità del modello finale;
- sono estremamente robusti rispetto agli *outliers*;
- l'output di una procedura ad albero è di facile comprensione ed interpretazione.

4.4.1 Contesto algoritmico

La diffusione degli alberi decisionali ha avuto inizio nel 1963 [Morgan and Sonquist, 1963]. La struttura degli alberi decisionali fa riferimento alla conformazione naturale degli alberi ed è simile ad un diagramma di flusso. Infatti, ogni nodo interno denota un test su una variabile, ogni ramo rappresenta un risultato del test mentre i nodi esterni, denominati foglie, rappresentano valori relativi alla variabile obiettivo oppure una distribuzione di classi (Figura 4.13).

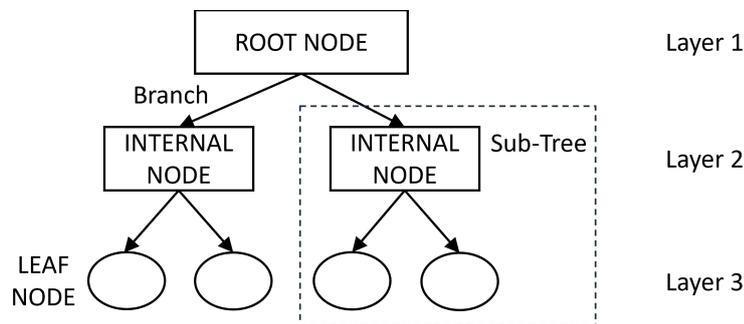


Figura 4.13: Schematizzazione della struttura di un albero decisionale.

La generazione di un albero decisionale avviene sulla base dei dati di addestramento in una direzione dall'alto verso il basso, ovvero dal generale allo specifico. Lo stadio iniziale di un albero decisionale è il nodo radice, al quale vengono assegnati tutti i punti dell'insieme di addestramento. Se tutti i punti hanno la stessa uscita (regressione) oppure appartengono alla stessa classe (classificazione), allora non sarà più necessario prendere ulteriori decisioni in merito alla suddivisione dei dati e la soluzione sarà completa. Tuttavia, se i punti in questo nodo hanno diverse uscite (regressione) oppure appartengono a due o più classi (classificazione), allora sarà necessario effettuare un test sul nodo che darà luogo ad una divisione (*split*). Il processo viene ripetuto ricorsivamente per ciascuno dei nuovi nodi intermedi fino ad ottenere

un albero completamente discriminante. Un albero decisionale in questa fase è potenzialmente una soluzione sovra-adattata, cioè può avere componenti troppo specifiche per il rumore e gli *outliers* che possono essere presenti nei dati di addestramento. Per attenuare questo *overfitting*, la maggior parte dei metodi ad albero decisionale attraversa una seconda fase, chiamata potatura (*pruning*), che cerca di generalizzare l'albero eliminando i sotto-alberi che risultano troppo specifici. Le tecniche di stima degli errori svolgono un ruolo fondamentale nella fase di potatura degli alberi. Infatti, la maggior parte dei moderni algoritmi di modellazione degli alberi decisionali consiste in una combinazione di un criterio specifico di divisione (per la completa crescita dell'albero) e di un criterio specifico di potatura per l'eliminazione dei sotto-alberi troppo specifici.

Il meccanismo di funzionamento di un DT è molto trasparente ed, osservando la sua struttura, è semplice capire in che modo viene presa ciascuna decisione. Questo fattore spiega come i DTs siano una delle tecniche maggiormente impiegate nel contesto dell'apprendimento automatico. Nel caso di un albero decisionale binario, ogni nodo interno ha esattamente due figli, quindi una decisione può sempre essere interpretata come vero o falso. Di tutti gli alberi decisionali, quelli binari sono i più utilizzati per la loro semplicità e per l'ampia conoscenza delle loro caratteristiche. I DTs utilizzati per i problemi di classificazione sono spesso chiamati alberi di classificazione e ogni nodo terminale contiene un'etichetta che indica la classe prevista per un dato vettore di input. Allo stesso modo, i DTs utilizzati per i problemi di regressione sono chiamati alberi di regressione e le etichette dei nodi terminali possono essere delle costanti oppure delle equazioni che specificano il valore di uscita previsto per un dato vettore di ingresso [Jang et al., 1997]. La Figura 4.14 rappresenta una tipica struttura di albero decisionale binario con due variabili di input x_1 e x_2 ed un unico output. Tale albero decisionale partiziona lo spazio degli input in quattro aree rettangolari non sovrapposte, a ciascuna delle quali viene assegnata un'etichetta A_i per rappresentare un valore di output previsto.

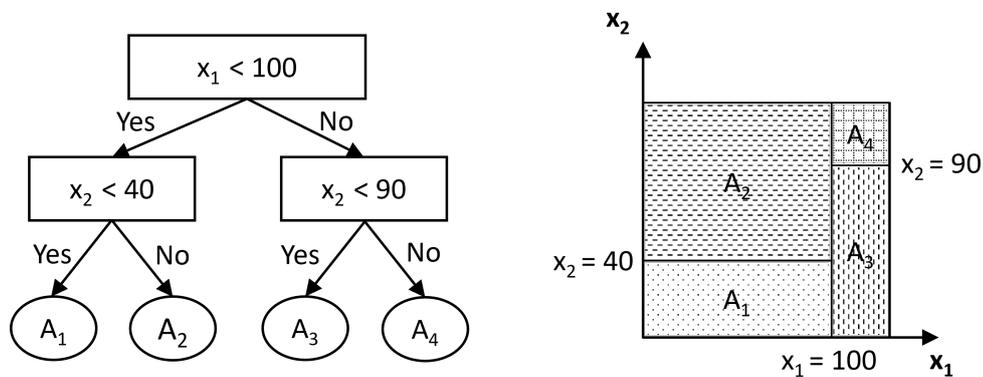


Figura 4.14: Esempio di albero decisionale binario con due variabili di input.

I principali algoritmi di DTs presenti all'interno della letteratura scientifica sono ID3 [Quinlan, 1986], C4.5 [Quinlan, 1993] e CART [Breiman et al., 1984]. Tuttavia, per ragioni legate alla semplificazione della trattazione, una descrizione di maggior dettaglio verrà fornita esclusivamente per l'algoritmo CART.

4.4.2 Criteri di divisione

Nella maggior parte dei casi, le funzioni di suddivisione discrete sono univariate. Ciò significa che un nodo interno viene suddiviso in base al valore di una singola variabile di ingresso. Di conseguenza, il sistema induttore cerca la variabile migliore su cui effettuare la suddivisione. I criteri univariati possono essere caratterizzati in diversi modi, quali: (i) in base all'origine della misura (teoria dell'informazione, dipendenza e distanza); (ii) in base alla struttura della misura: criteri basati sulle impurità, criteri basati sulle impurità normalizzate e criteri binari. I criteri più comuni presenti in letteratura sono basati su metriche note come

- Impurità: assegnata una variabile casuale x con k valori discreti distribuiti secondo $P = (p_1, p_2, \dots, p_k)$, l'impurità è definita come una funzione $\phi : [0, 1]^k \rightarrow R$ che soddisfi le seguenti condizioni:
 - $\phi(P) \geq 0$;
 - $\phi(P)$ è minimo se $\exists i$ tale che $p_i = 1$;
 - $\phi(P)$ è massimo se $\forall i, 1 \leq i \leq k, p_i = \frac{1}{k}$;
 - $\phi(P)$ è simmetrico rispetto alle componenti di P ;
 - $\phi(P)$ è regolare (sempre differenziabile) nel suo intervallo;

Si noti che nel caso in cui il vettore delle probabilità P abbia una componente pari a 1 (ovvero che la variabile x assuma un unico valore), in tal caso x sarà definita pura. D'altra parte, nel caso in cui tutte le componenti di P siano uguali, il livello di impurità raggiungerà il valore massimo. Assegnato un insieme di osservazioni di addestramento S , il vettore delle probabilità dell'attributo target y è definito come [Maimon and Rokach, 2005]:

$$P_y(S) = \left(\frac{|\sigma_{y=c_1} S|}{|S|}, \dots, \frac{|\sigma_{y=c_{|dom(y)|}} S|}{|S|} \right) \quad (4.4.1)$$

La bontà della divisione dovuta all'attributo discreto a_i è definita come la riduzione dell'impurità dell'attributo target in seguito alla suddivisione di S in base ai valori $v_{i,j} \in dom(a_i)$:

$$\Delta\Phi(a_i, S) = \phi(P_y(S)) - \sum_{j=1}^{|dom(a_i)|} \frac{|\sigma_{a_i=v_{i,j}} S|}{|S|} \cdot \phi(P_y(\sigma_{a_i=v_{i,j}} S)) \quad (4.4.2)$$

- Guadagno di informazione (*Information Gain* - IG), utilizza la misura dell'entropia (derivata dalla teoria dell'informazione) come misura dell'impurità [Quinlan, 1986]:

$$IG(a_i, S) = Entropia(y, S) - \sum_{v_{i,j} \in dom(a_i)} \frac{|\sigma_{a_i=v_{i,j}} S|}{|S|} \cdot Entropia(y, \sigma_{a_i=v_{i,j}} S) \quad (4.4.3)$$

dove:

$$Entropia(y, S) = \sum_{c_j \in dom(y)} - \frac{|\sigma_{y=c_j} S|}{|S|} \cdot \log_2 \frac{|\sigma_{y=c_j} S|}{|S|} \quad (4.4.4)$$

- Indice di Gini, misura le differenze tra le distribuzioni di probabilità dei valori dell'attributo target. L'indice di Gini è stato utilizzato in diversi lavori come [Breiman et al., 1984] e [Gelfand et al., 1989] ed è definito come:

$$\text{Gini}(y, S) = 1 - \sum_{c_j \in \text{dom}(y)} \left(\frac{|\sigma_{y=c_j} S|}{|S|} \right)^2 \quad (4.4.5)$$

Di conseguenza, il criterio di valutazione per la selezione dell'attributo a_i si definisce come:

$$\text{Gini Gain}(a_i, S) = \text{Gini}(y, S) - \sum_{v_{ij} \in \text{dom}(a_i)} \frac{|\sigma_{a_i=v_{ij}} S|}{|S|} \cdot \text{Gini}(y, \sigma_{a_i=v_{ij}} S) \quad (4.4.6)$$

4.4.3 Criteri di arresto

La fase di crescita di un albero decisionale continua finché non viene innescato un criterio di arresto. I criteri di arresto più comuni sono di seguito elencati [Maimon and Rokach, 2005]:

- tutte le istanze del set di addestramento appartengono ad un unico valore di y ;
- è stata raggiunta la massima profondità dell'albero;
- il numero di casi del nodo terminale è inferiore al numero minimo di casi per i nodi genitori;
- se i nodi fossero divisi, allora il numero di casi in uno o più nodi figli sarebbe inferiore al numero minimo di casi per i nodi figli;
- il miglior criterio di divisione non è maggiore di una determinata soglia;

4.4.4 Metodi di potatura

L'impiego di criteri di arresto eccessivamente restrittivi tende ad originare alberi decisionali piccoli e poco adattati. D'altra parte, l'utilizzo di criteri di arresto eccessivamente flessibili tende a generare alberi decisionali di grandi dimensioni che si sovra-adattano all'insieme di dati di addestramento. Per risolvere questo problema sono stati sviluppati i metodi di potatura suggeriti originariamente da Breiman [Breiman et al., 1984]. Secondo questa filosofia, si utilizza un criterio di arresto molto flessibile, lasciando che l'albero decisionale si sovra-adatti all'insieme di dati di addestramento. Successivamente l'albero sovradimensionato viene ridotto in un albero più piccolo, eliminando i sotto-alberi che non contribuiscono all'accuratezza della generalizzazione. I benefici apportati dall'utilizzo dei metodi di potatura sono osservabili soprattutto in caso di domini molto rumorosi. Un'altra motivazione chiave della potatura è "negoziare l'accuratezza per la semplicità" [Bohanec and Bratko, 1994]. Quando l'obiettivo è produrre una descrizione del concetto compatta e sufficientemente accurata, la potatura è molto utile. All'interno di questo processo, l'albero decisionale iniziale viene visto come un albero perfettamente accurato. Pertanto, l'accuratezza di un albero decisionale potato

indica quanto questo sia vicino all'albero iniziale [Maimon and Rokach, 2005]. Esistono diverse tecniche di potatura degli alberi decisionali. In linea generale, un nodo viene potato se questa operazione apporta miglioramenti in funzione di un determinato criterio. Una delle tecniche maggiormente utilizzate prende il nome di *Cost Complexity Pruning* (CCP). Si svolge in due fasi [Breiman et al., 1984]. In un primo momento si costruisce una sequenza di alberi T_0, T_1, \dots, T_k sulla base dell'insieme dei dati di addestramento, con T_0 l'albero originale prima della potatura e T_k l'albero radice. In un secondo momento, uno di questi k alberi viene scelto come albero potato, in base alla stima del suo errore di generalizzazione. L'albero T_{i+1} è ottenuto sostituendo uno o più sottoalberi dell'albero precedente T_i con delle foglie adeguate. I sottoalberi potati sono quelli che restituiscono l'aumento più basso del tasso di errore apparente per ciascuna foglia potata:

$$\alpha = \frac{\epsilon(\text{potato}(T, t), S) - \epsilon(T, S)}{|foglie(T)| - |foglie(\text{potato}(T, t))|} \quad (4.4.7)$$

dove $\epsilon(T, S)$ rappresenta il tasso di errore dell'albero T sulla base del campione S , $|foglie(T)|$ rappresenta il numero di foglie in T , $\text{potato}(T, t)$ rappresenta l'albero ottenuto sostituendo il nodo t in T con un'adeguata foglia.

In una seconda fase viene stimato l'errore di generalizzazione di ciascun albero potato T_0, T_1, \dots, T_k . Il miglior albero potato viene dunque selezionato. Ammesso che la banca dati sia sufficientemente estesa, è consigliabile suddividerlo in un insieme di addestramento ed uno di potatura in modo tale che l'albero venga costruito utilizzando l'insieme di addestramento e valutato su quello di potatura. Se questa ipotesi non dovesse essere soddisfatta, è alternativamente possibile utilizzare una metodologia di convalida incrociata descritta al paragrafo 3.6.

4.4.5 Algoritmo CART

L'algoritmo CART (*Classification and Regression Tree*) è stato presentato per la prima volta nel 1984 [Breiman et al., 1984]. Esso consiste nella generazione di un albero che dapprima cresce estensivamente sulla base dell'insieme di osservazioni di addestramento e poi viene potato sulla base del metodo del minimo CCP (Equazione 4.4.7). In linea generale, il CART suddivide le osservazioni in due sottoinsiemi in modo tale che, progressivamente, ciascun sottoinsieme risulti il più omogeneo e puro possibile. Si tratta dunque di un processo iterativo che termina quando si raggiunge un predeterminato criterio di omogeneità oppure quando si innesca un particolare criterio d'arresto. Il CART misura l'impurità di una divisione in un nodo definendo una metrica di impurità ed una stessa variabile di input ha la possibilità di essere utilizzata più volte a diversi livelli dell'albero. Per ciascuna divisione dell'albero, il CART identifica le variabili di input più simili alla variabile di divisione selezionata. Tali variabili sono definite surrogate per quella specifica divisione. Ad esempio, quando un'osservazione dev'essere classificata ma presenta un valore mancante per la variabile di divisione, il suo valore in termini della variabile surrogata può essere utilizzato per effettuare la divisione stessa. Questa impostazione consente una maggiore flessibilità nella gestione dei valori mancanti, al prezzo di un onere computazionale maggiore e tempi di addestramento più lunghi.

Il CART utilizza l'indice di Gini come criterio di divisione (Equazione 4.4.6). Per quanto riguarda le potenziali divisioni, il CART considera tutte le possibilità sia per le variabili continue ($n - 1$ potenziali divisioni per n valori distinti) che per quelle categoriche ($2^n - 1$ potenziali divisioni per n valori distinti). Per ciascuna potenziale divisione, viene sommata l'impurità (definita dall'indice di Gini) di tutte le sotto-partizioni e viene applicata la sola divisione che comporta la massima riduzione di impurità.

4.4.6 Modelli Ensemble

Gli alberi decisionali funzionano molto bene in svariate applicazioni, tuttavia esistono delle metodologie che permettono di raggiungere performance migliori assemblando molteplici architetture ad alberi. Queste tecniche prendono il nome di metodi *ensemble* ed offrono prestazioni generalmente superiori al prezzo di una complessità computazionale ed algoritmica maggiori. Nei modelli *ensemble*, il singolo albero decisionale prende il nome di *weak learner* (o modello semplice) e viene solitamente combinato con altri *weak learners* al fine di generare un modello complessivo solido e robusto. Ogni *weak learner* delle metodologie *ensemble* coglie alcuni aspetti dell'informazione contenuta nei dati utilizzati per l'addestramento. Il compito dell'*ensemble tree* è quello di combinare in maniera ottimale i *weak learners* al fine di ottenere metriche di performance complessivamente migliori. Il vantaggio principale dei metodi ensemble è la riduzione del fenomeno dell'*overfitting*. Esistono tre principali tecnologie che permettono di generare modelli *ensemble*:

- *Stacking*, in cui l'idea è addestrare molteplici modelli semplici (anche diversi tra loro) e combinarli all'interno di un meta-modello in grado di produrre previsioni più accurate basate sulle previsioni multiple degli *weak learners*. Pertanto, ai fini dell'applicazione di una tecnica di *stacking*, sarà necessario definire gli L *weak learners* da addestrare ed il meta-modello che li combini;
- *Bagging*, in cui l'obiettivo è ridurre la varianza di un albero decisionale. L'idea di partenza è quella di generare in modo casuale e per sostituzione diversi sottoinsiemi di dati a partire dall'insieme di addestramento. Ciò significa che ciascun valore/istanza viene ripetuto/a due volte (o più) all'interno di un campione. Ogni sottoinsieme viene dunque utilizzato per addestrare in parallelo ed in maniera indipendente il proprio albero decisionale. Infine, a seconda dell'attività (regressione o classificazione), vengono rispettivamente ricavate una media o una maggioranza delle previsioni al fine di ottenere una stima più accurata ed attendibile. In linea generale, avendo indicato con s_L il modello *ensemble* e ciascun *weak learner* con w_l , per la risoluzione di problemi di regressione è possibile scrivere:

$$s_L(\cdot) = \frac{1}{L} \sum_{l=1}^L w_l(\cdot) \quad (4.4.8)$$

- *Boosting*, in cui l'obiettivo è ridurre la distorsione dei modelli semplici che compongono il modello robusto. L'idea generale è quella di addestrare in maniera sequenziale i diversi modelli semplici, dando progressivamente maggiore importanza alle osservazioni

nel dataset che sono state mal gestite dai modelli semplici precedenti. Intuitivamente, ogni nuovo modello semplice concentra i propri sforzi sulle osservazioni maggiormente complesse da gestire e da apprendere generando un modello *ensemble* robusto. Essendo principalmente focalizzata sulla riduzione del bias, la tecnica *boosting* predilige quei modelli di base con bassa varianza ma ad elevata distorsione. Una volta scelti i *weak learners*, sarà necessario definire come questi saranno sequenzialmente addestrati ed aggregati. I principali algoritmi di *boosting* prendono il nome di *AdaBoost* e *Gradient Boosting*. Questi due meta-algoritmi differiscono nella maniera in cui creano ed aggregano i *weak learners* durante il processo sequenziale. *AdaBoost* aggiorna i pesi associati ad ogni osservazione del dataset di addestramento, mentre il *Gradient Boosting* aggiorna il valore di queste osservazioni. Questa differenza sostanziale deriva dal modo in cui entrambe queste metodologie cercano di risolvere il problema dell'ottimizzazione per identificare il miglior modello derivante dalla somma pesata di *weak learners*.

Per motivi legati alla semplificazione della trattazione, una descrizione più esaustiva sarà fornita esclusivamente per la metodologia *Gradient Boosting*.

4.4.7 Gradient Boosting

Il *Gradient Boosting* (GB) viene descritto da Jerome H. Friedman nello studio intitolato "Greedy function approximation: a gradient boosting machine" [Friedman, 2001]. Essendo una tecnica di ML supervisionata, la trattazione analitica partirà da un insieme $\{\mathbf{x}_i, y_i\}$ composto da valori di input \mathbf{x}_i e da valori attesi di output y_i , con $i \in \{1, \dots, n\}$. L'approccio GB consiste nel costruire iterativamente una serie di funzioni $F^0, F^1, \dots, F^t, \dots, F^m$, assegnata una funzione di perdita $L(y_i, F^t)$. Come si può notare, la funzione di perdita necessita di due valori di input: l' i -esimo valore atteso di output y_i e la t -esima funzione di perdita F^t che stimi y_i . Supponendo di aver costruito la funzione F^t , le stime di y_i possono essere migliorate attraverso una nuova funzione $F^{t+1} = F^t + h^{t+1}(\mathbf{x})$ tale che h^{t+1} minimizzi il valore atteso della funzione di perdita (*expected loss* - $\mathbb{E}\mathcal{L}$):

$$h^{t+1} = \arg \min_{h \in H} \mathbb{E}\mathcal{L}(y, F^t) \quad (4.4.9)$$

avendo indicato con H l'insieme dei candidati alberi decisionali tra cui scegliere quello da aggiungere al modello *ensemble*. Sulla base della definizione F^{t+1} è possibile scrivere il valore atteso della funzione di perdita \mathcal{L} in termini di F^t e h^{t+1} :

$$\mathbb{E}\mathcal{L}(y, F^{t+1}) = \mathbb{E}\mathcal{L}(y, F^t + h^{t+1}) \quad (4.4.10)$$

È possibile osservare che il lato destro dell'equazione 4.4.10 implica il tentativo di minimizzare il valore della funzione di perdita \mathcal{L} valutato sulla base di y ed F^t più *qualcosa*. Ipotizzando che \mathcal{L} sia continua e differenziabile, è possibile aggiungere a F^t qualcosa di relativo al tasso di variazione di \mathcal{L} per spostarne il valore nella direzione in cui \mathcal{L} sia decrescente. Pertanto, impostando h^{t+1} su valori nella direzione in cui il gradiente di \mathcal{L} rispetto a F^t diminuisce più rapidamente, si otterrà il valore h^{t+1} che approssimativamente minimizza $\mathbb{E}\mathcal{L}(y, F^t + h^{t+1})$. Sulla base di queste premesse, è possibile scrivere un'approssimazione ragionevole per h^{t+1} :

$$h^{t+1} \approx \arg \min_{h \in H} \mathbb{E} \left(\frac{\partial \mathcal{L}y}{\partial F^t} - h \right)^2 \quad (4.4.11)$$

L'utilizzo delle derivate parziali della funzione di perdita \mathcal{L} rispetto alla funzione F^t per trovare h^{t+1} è ciò che conferisce a questa tecnica di *boosting* il termine *gradient*. Prokhorenkova et al., però, sottolineano che potrebbe non essere semplice calcolare la quantità $\arg \min_{h \in H} \mathbb{E} \left(\frac{\partial \mathcal{L}y}{\partial F^t} - h \right)^2$. Questo in virtù del fatto che potrebbe essere complesso definire la probabilità di ottenere specifici valori della quantità $\arg \min_{h \in H} \mathbb{E} \left(\frac{\partial \mathcal{L}y}{\partial F^t} - h \right)^2$ [Prokhorenkova et al., 2018]. Inoltre, l'ipotesi di utilizzare tecniche stocastiche per la definizione di F^t (come ad esempio alcuni algoritmi utilizzati per costruire alberi decisionali) potrebbe complicare ulteriormente la conoscenza di F^t . Per questo motivo, Prokhorenkova et al. suggeriscono l'assunzione:

$$\arg \min_{h \in H} \mathbb{E} \left(\frac{\partial \mathcal{L}y}{\partial F^t} - h \right)^2 \approx \arg \min_{h \in H} \frac{1}{n} \left(\frac{\partial \mathcal{L}y}{\partial F^t} - h \right)^2 \quad (4.4.12)$$

Utilizzando le approssimazioni fornite nelle equazioni 4.4.11 e 4.4.12, è possibile ottenere una valida stima di h^{t+1} :

$$h^{t+1} \approx \arg \min_{h \in H} \frac{1}{n} \left(\frac{\partial \mathcal{L}y}{\partial F^t} - h \right)^2 \quad (4.4.13)$$

Nel caso di *Gradient Boosting Decision Tree* (GBDT), il modello base F^0 è un albero decisionale così come sono alberi decisionali anche $h^1, h^2, \dots, h^t, \dots, h^m$. Aggiungendo un DT per costruire F^{t+1} secondo le modalità descritte, il valore stimato della funzione di perdita $\mathbb{E} \mathcal{L}(y, f^{t+1}(\mathbf{x}))$ si riduce, implicando che le stime $F^{j+1}(\mathbf{x}_i)$ siano migliori di quelle $F^j(\mathbf{x}_i)$.

Attualmente, molteplici tecniche GBDT apportano miglioramenti alla classica tecnica GB descritta da Friedman in [Friedman, 2001]. Queste prendono il nome di *eXtreme Gradient Boosting* (XGBoost) [Chen and Guestrin, 2016], *Light Gradient Boosting Method* (LightGBM) [Ke et al., 2017] e *Categorical Boosting* (CatBoost) [Prokhorenkova et al., 2018]. Tuttavia, per ragioni legate alla semplificazione della trattazione, una descrizione di maggior dettaglio verrà fornita esclusivamente per l'algoritmo CatBoost.

4.4.8 CatBoost

Sono molteplici i perfezionamenti apportati dall'algoritmo CatBoost al tradizionale *Gradient Boosting*: in prima battuta, è necessario considerare il modo in cui CatBoost gestisce le variabili categoriche ad elevata cardinalità. Infatti, per le variabili categoriche a bassa cardinalità, CatBoost utilizza la tradizionale codifica *one-hot* (in cui tutti i bit di un gruppo sono posti pari a 0 tranne uno posto pari a 1). Per la codifica delle variabili categoriche ad elevata cardinalità, invece, il CatBoost utilizza una metodologia nota come *Ordered Target Statistic* (OTS). Per statistica target si intende un valore calcolato a partire dai veri valori di output associati a particolari valori della variabile categorica di una banca dati. Pertanto, una strategia per gestire le variabili categoriche ad elevata cardinalità, consiste nel sostituire i valori categorici di una variabile con una statistica target.

Il concetto più importante nel calcolo dell'OTS deriva dalla distinzione tra insieme di osservazioni di addestramento e di test. Sia \mathcal{D} l'insieme di tutti i dati disponibili per addestrare e valutare il modello *ensemble*. L'albero decisionale h^{t+1} aggiunto all'*ensemble tree* minimizza il valore atteso della funzione di perdita $\mathbb{E}\mathcal{L}$. L'obiettivo è utilizzare alcuni dati presenti in \mathcal{D} per addestrare il DT h^{t+1} ed altri per identificare l' h^{t+1} che minimizzi $\mathbb{E}\mathcal{L}$. La motivazione principale perché i dati in \mathcal{D} vengano utilizzati per tale scopo deriva dal fatto che si cerchi di evitare quel fenomeno che in letteratura [Prokhorenkova et al., 2018] viene descritto come *target leakage* e che sarà successivamente descritto nel dettaglio. La modalità con cui CatBoost seleziona i dati da utilizzare per l'addestramento del DT h^{t+1} consiste nel conferire un ordine arbitrario agli elementi di \mathcal{D} attraverso una permutazione casuale σ . Siano $\sigma(k)$ il k -esimo elemento di \mathcal{D} in base a σ e $\mathcal{D}_k = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k-1}\}$ ordinato sulla base della permutazione casuale σ . CatBoost utilizza \mathcal{D}_k come dati per l'addestramento del DT h^{t+1} e \mathcal{D} come dati per valutare se h^{t+1} sia un DT che minimizzi $\mathbb{E}\mathcal{L}(y, F^t + h^{t+1})$.

Un secondo concetto importante per comprendere la codifica CatBoost per la gestione delle variabili categoriche ad elevata cardinalità risiede nel funzionamento dell'operatore $\mathbb{1}$. L'operatore $\mathbb{1}_{a=b}$ è una funzione di una variabile a che risulta pari a 1 quando $a = b$, 0 altrimenti. Tale operatore gioca un ruolo cruciale nella formula che CatBoost applica per mappare i valori di una variabile categorica in valori numeri. Nello specifico, tale formula coinvolge l'utilizzo dell'operatore $\mathbb{1}_{x_j^i = x_k^i}$ che, pertanto, risulterà pari a 1 quando l' i -esima componente del vettore di input di CatBoost \mathbf{x}_j è uguale all' i -esima componente del vettore di input \mathbf{x}_k . In questo caso, l'indicatore k è quello relativo al k -esimo elemento secondo l'ordine impartito a \mathcal{D} in funzione della permutazione casuale σ , mentre i assume valori pari a numeri interi compresi tra 1 e $k - 1$.

La comprensione di questi due concetti chiave consente di definire la formula necessaria per ottenere il valore codificato, \hat{x}_k^i , dell' i -esima variabile categorica del k -esimo elemento di \mathcal{D} :

$$\hat{x}_k^i = \frac{\sum_{x_j \in \mathcal{D}_k} \mathbb{1}_{x_j^i = x_k^i} \cdot y_j + ap}{\sum_{x_j \in \mathcal{D}_k} \mathbb{1}_{x_j^i = x_k^i} + a} \quad (4.4.14)$$

Prokhorenkova et al. definiscono il parametro a come un valore positivo (maggiore di 0) ed il parametro p come il valore medio della variabile categorica all'interno della banca dati. Si può notare come l'impostazione $a > 0$ consenta di non dividere per 0 nel caso in cui nessuno dei valori x_j^i sia uguale a x_k^i . Inoltre, in questa particolare condizione, qualsiasi valore $a > 0$ garantisce che \hat{x}_k^i assuma il valore p . L'algoritmo CatBoost applica l'equazione 4.4.14 durante l'addestramento del DT h^{t+1} ma ne utilizza una variante durante la valutazione di h^{t+1} per determinare se sia il DT che minimizza $\mathbb{E}\mathcal{L}(y, F^t + h^{t+1})$. La variante consiste nell'utilizzare l'intero insieme \mathcal{D} piuttosto che un sottoinsieme \mathcal{D}_k . Come accennato in precedenza, la motivazione principale che ha spinto Prokhorenkova et al. ad utilizzare questo particolare tipologia di codifica è il tentativo di evitare il fenomeno del *target leakage*. Tale condizione risulta verificata in virtù del fatto che la codifica effettuata secondo l'equazione 4.4.14 goda di una particolare proprietà, ovvero:

$$\mathbb{E}(\hat{x}^i | y = v) = \mathbb{E}(\hat{x}_k^i | y_k = v) \quad (4.4.15)$$

Un aspetto interessante risiede nel fatto che il modo in cui la tecnica di codifica CatBoost soddisfa la proprietà 4.4.15 sia quello di assicurare che il valore y_k non venga utilizzato nell'equazione 4.4.14. Infatti, Prokhorenkova et al. hanno dimostrato che utilizzando y_k per codificare le variabili \mathbf{x}_k si dà origine all' indesiderato fenomeno del *target leakage*. Se l'equazione 4.4.15 non dovesse risultare verificata, significherebbe che il valore atteso di tutti i valori codificati per l' i -esima variabile (assegnato uno specifico valore di output v) non coincidano con il valore atteso dei valori codificati per alcuni esempi di addestramento (\mathbf{x}_k, y_k) . In altre parole, si tratterebbe di una particolare condizione di *overfitting* per la quale il modello è in grado di sfruttare la correlazione tra \hat{x}_k e y_k durante l'addestramento ma, allo stesso tempo, tale correlazione non sussisterà durante il test a causa della differenza nei valori attesi. La soluzione suggerita perché si eviti lo scostamento dei valori attesi secondo le condizioni $y = v$ e $y_k = v$ consiste nell'escludere il valore di y_k dal calcolo dei valori di \hat{x}_k^i durante la codifica del valore x_k^i ; da questa considerazione nasce la necessità di definire \mathcal{D}_k così come suggerito nell'equazione 4.4.14.

Una seconda importante proprietà dell'OTS descritta da Prokhorenkova et al. è che questa tecnica permette di utilizzare tutti gli esempi di addestramento (\mathbf{x}_k, y_k) . Infatti, al termine di un numero sufficiente di iterazioni, tutti i valori categorici saranno stati codificati con tutte le informazioni disponibili all'interno dei dati di addestramento. Questa seconda proprietà bilancia la prevenzione dell'*overfitting* della prima proprietà, garantendo che non si verifichi il fenomeno altrettanto indesiderato dell'*underfitting*. La maniera in cui Prokhorenkova et al. sfruttano questa proprietà costituisce un'ulteriore ottimizzazione del Gradient Boosting che gli autori denominano *Ordered Boosting* (OB). Il *target leakage* provoca non solo uno scostamento condizionale nel valore atteso di una variabile codificata, ma anche uno scostamento previsionale nel valore atteso dei residui che si cerca di minimizzare. Per poter apprezzare quest'ultima osservazione, si consideri l'equazione 4.4.13 e si assuma di utilizzare la tecnica OTS per codificare alcune variabili categoriche e costruire il GBDT che generi F^{t+1} . Con queste premesse, $\partial\mathcal{L}y/\partial F^t$ è una variabile casuale poiché una permutazione casuale $\sigma(k)$ è stata utilizzata al fine di scegliere gli elementi di \mathcal{D}_k per codificare le variabili categoriche che influenzano il valore di F^t . Pertanto, la distribuzione di $\partial\mathcal{L}y/\partial F^t$ è stata calcolata per una particolare codifica x_k^i . Questo scostamento condizionale provoca un *bias* nella stima di h^{t+1} che impatta negativamente sulle performance ottenute valutando F^{t+1} su dati non utilizzati in fase di addestramento (ovvero sulle capacità di generalizzazione). La tecnica dell'*Ordered Boosting* torna utile proprio per contrastare tale fenomeno. Il concetto chiave nell'OB risiede nell'utilizzare gli stessi esempi di \mathcal{D}_k utilizzati per determinare OTS anche per calcolare le stime di h^{t+1} , ovvero per computare i valori di $\partial\mathcal{L}y/\partial F^t$. Come specificato in precedenza, $\mathcal{D}_k = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k-1}\}$ dipende dal punto in cui ci si trova nell'iterazione fra le permutazioni degli elementi di \mathcal{D} . In altre parole, con $k = 1$, \mathcal{D}_k sarà composto da un solo elemento. Ciò significa che la varianza dei valori stimati di $\partial\mathcal{L}y/\partial F^t$ sarà particolarmente alta. Pertanto, durante l'OB, CatBoost utilizza molteplici ed indipendenti permutazioni $\sigma_1, \sigma_2, \dots, \sigma_s$ di \mathcal{D} per calcolare un certo numero di insiemi di valori residui da utilizzare per identificare h^{t+1} (per ottenere F^{t+1}) e garantire che nessuno dei valori di x_k^i sia impiegato per calcolare i valori dei gradienti $\partial\mathcal{L}y/\partial F^t$. Al contempo, utilizzare questi molteplici insiemi di residui riduce la varianza nelle stime di $\partial\mathcal{L}y/\partial F^t$ effettuate dall'algoritmo CatBoost.

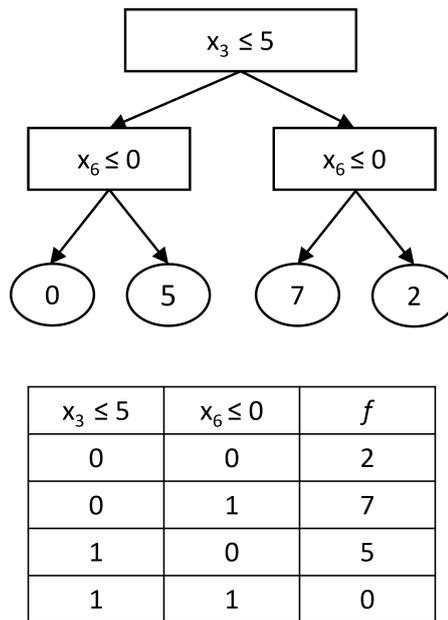


Figura 4.15: Esempio di albero decisionale obliquo tratto da [Lou and Obukhov, 2017].

Infine, un ultimo concetto fondamentale nel processo di costruzione degli alberi decisionali tramite algoritmo CatBoost è quello degli alberi decisionali obliqui (*Oblivious Decision Trees* - ODT), chiamate anche tabelle decisionali. CatBoost, infatti, costruisce un insieme di ODT. Gli ODT sono alberi binari completi; pertanto, nel caso in cui l'ODT possieda n livelli, avrà conseguentemente 2^n nodi. Inoltre, ad esclusione delle foglie, tutti i nodi dell'ODT saranno dotati dello stesso criterio di suddivisione. Secondo Prokhorenkova et al., gli ODT sono bilanciati, meno inclini all'*overfitting* e consentono di accelerare notevolmente l'esecuzione al momento del test [Prokhorenkova et al., 2018]. Un esempio di albero decisionale obliquo (o tabella decisionale) viene mostrato in Figura 4.15, tratto da [Lou and Obukhov, 2017].

Gli ODT sono bilanciati per definizione. Poiché sono alberi binari completi, il numero di valutazioni per raggiungere un nodo foglia è rappresentato dal numero minimo di valutazioni necessarie per raggiungere il numero massimo di nodi foglia. Per questo motivo, gli ODT sono in grado di produrre esecuzioni più efficienti rispetto ad alberi decisionali più profondi ma incompleti. Il compromesso è rappresentato dal fatto che bisogna prestare molta attenzione all'impostazione della profondità massima dell'albero CatBoost, poiché la quantità di memoria utilizzata da CatBoost può crescere di un fattore pari a 2 volte il numero degli alberi nel modello *ensemble* per ogni aumento unitario della profondità massima dell'albero. Questo è un altro esempio chiave che dimostri la sensibilità di CatBoost alle impostazioni degli iperparametri, di cui i ricercatori dovrebbero essere pienamente consapevoli, poiché tale impostazione esercita un impatto non trascurabile sulla quantità di memoria e sul tempo di esecuzione delle simulazioni.

Capitolo 5

Approccio di Ricerca

” *Be less curious about people and more curious about ideas.*

– Marie Curie

Questo breve capitolo spiega *step-by-step* l’approccio utilizzato per raggiungere gli scopi di questa tesi, articolandolo nelle diverse fasi successive. Queste possono essere sintetizzate e schematizzate (Figura 5.1) come segue:

- Definizione del fenomeno da investigare e modellare: nei capitoli successivi verranno analizzati quattro problemi molto diversi tra loro, come ad esempio lo studio della capacità portante di una pista di volo, la caratterizzazione prestazionale di conglomerati bituminosi attraverso la stabilità Marshall ed altri parametri, la valutazione delle caratteristiche volumetriche e del comportamento meccanico di miscele contenenti filler sostenibili e l’analisi del modulo complesso di conglomerati bituminosi;
- Raccolta dei dati: sulla base dei rispettivi protocolli e delle corrispondenti normative, una quantità sufficiente di dati riferiti alle molteplici variabili coinvolte nel fenomeno da investigare è stata registrata e raccolta all’interno di banche dati che saranno successivamente utilizzate per la modellazione predittiva;
- Preparazione della banca dati: i dati raccolti nella fase precedente sono stati puliti, preparati, normalizzati ed organizzati selezionando le variabili che costituiranno gli input dei diversi modelli predittivi e quelle che invece saranno restituite come output;
- Scelta dell’algoritmo di regressione: sulla base dell’esperienza del *data analyst* e del problema in esame, è opportuno selezionare la tecnica di regressione ritenuta più opportuna per modellare il fenomeno in esame. Nella maggior parte dei casi, risulta conveniente selezionare molteplici algoritmi e successivamente confrontare le prestazioni raggiunte da ciascuna metodologia, in modo tale da ottenere una più ampia consapevolezza delle effettive capacità predittive dei modelli. Le metodologie adottate nell’ambito di questo elaborato hanno coinvolto diverse tecniche di regressione lineare, le macchine a vettori di supporto, nonché le reti neurali artificiali e gli alberi decisionali; tuttavia, in virtù di

una maggiore potenza computazionale e di risultati caratterizzati da migliore precisione predittiva [Rondinella et al., 2023], le applicazioni successive si limiteranno a descrivere i risultati ottenuti utilizzando le sole reti neurali (sia superficiali che profonde) e gli alberi decisionali sviluppati mediante l'innovativo algoritmo CatBoost;

- Ottimizzazione degli iperparametri del modello: al fine di identificare il set di iperparametri che restituisce le migliori performance predittive di ciascun modello sviluppato, molteplici procedure sono state seguite tra cui la ricerca a griglia o l'ottimizzazione Bayesiana accoppiate con la convalida incrociata;
- Valutazione delle performance: la bontà di adattamento di ciascun modello ai dati a disposizione è stata infine valutata mediante molteplici metriche di regressione come ad esempio MAE, MAPE, MSE, RMSE, R e R^2 .

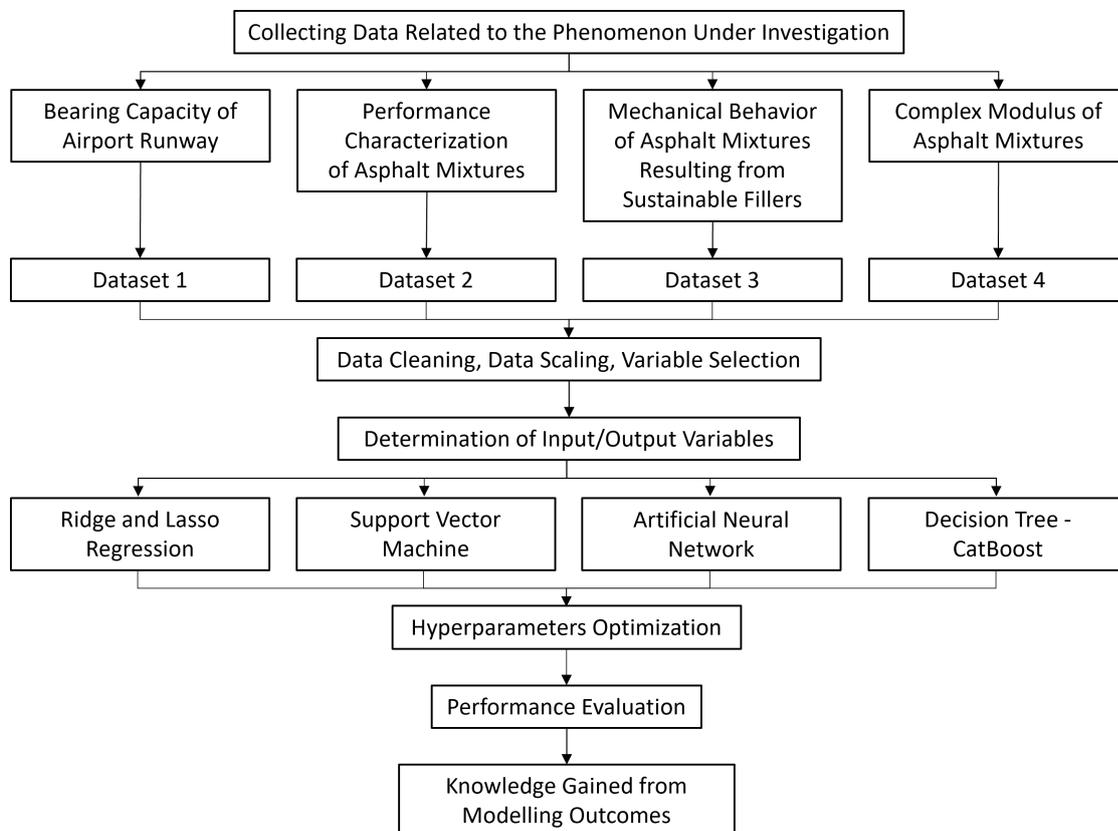


Figura 5.1: Schematizzazione approccio di ricerca.

Capitolo 6

Capacità Portante di una Pista di Volo

” *An airport runway is the most important Main Street in any town.*

— Norman Crabtree

6.1 Introduzione

Le reti stradali e le piste di volo aeroportuali rappresentano degli *asset* fondamentali sia per i Paesi sviluppati che per quelli in via di sviluppo [Schweikert et al., 2014]. I costi relativi alle operazioni di manutenzione rappresentano le voci di spesa più significative per i Paesi sviluppati, mentre quelli associati ai lavori di prima realizzazione sono più rilevanti per i Paesi in via di sviluppo [Makovšek et al., 2012]. Tuttavia, sia la prima costruzione che la manutenzione richiedono un notevole consumo di risorse naturali in termini di aggregati e legante bituminoso. Per questo motivo, tali infrastrutture di trasporto hanno un notevole impatto sia sul consumo di energia per le coinvolte produzioni industriali sia sulle relative emissioni, poiché la loro vita utile è fissata in diversi decenni [Garg et al., 2004]. In questo contesto, la previsione delle prestazioni a lungo termine di queste infrastrutture è fondamentale al fine di implementare correttamente le dovute strategie di manutenzione e riabilitazione (*Maintenance and Rehabilitation - M&R*) ed ottenere soluzioni tecnologiche economiche e sostenibili dal punto di vista ambientale [Ashtiani et al., 2021]. È ampiamente riconosciuto che una gestione efficiente di un aeroporto abbia diversi effetti ambientali positivi, come ad esempio: (a) la riduzione della congestione del traffico aeroportuale; (b) la minimizzazione del consumo di terreno ottimizzando lo sfruttamento delle infrastrutture esistenti [Carlucci et al., 2018]; (c) il miglioramento complessivo della sostenibilità delle costruzioni aeroportuali. In generale, la pista di volo rappresenta l'infrastruttura aeroportuale più importante in quanto deve garantire un'adeguata operatività in qualsiasi condizione meteorologica e/o di traffico. Tuttavia, esistono molte cause che possono portare al suo deterioramento, come il progressivo invecchiamento, la crescente domanda di traffico e l'assenza di investimenti adeguati [Rix et al., 1995]. Pertanto, l'utilizzo di tecniche di indagine sempre più innovative ed affidabili [Zou et al., 2020] si rende necessario per valutarne l'integrità strutturale, rispettando i limiti imposti dai vincoli di bilancio [Guo et al., 2016].

Le valutazioni strutturali delle piste di volo forniscono preziose informazioni sul corrispondente comportamento atteso [Haas et al., 1994] e sono estremamente utili per stimarne l'attuale idoneità tecnica e la vita utile residua. Inoltre, aiutano a definire le strategie di riabilitazione da adottare, nell'ambito di un adeguato sistema di gestione delle pavimentazioni aeroportuali (*Airport Pavement Management System* - APMS). Un APMS si basa sul concetto di manutenzione preventiva [Zheng et al., 2021, De Souza and De Almeida Filho, 2020], ossia intervenire prima che si verifichino condizioni di emergenza e la funzionalità della pavimentazione risulti compromessa. Lo scopo principale di un APMS è assistere l'Autorità di Gestione nell'individuare strategie affidabili ed economiche per preservare l'adeguatezza e la conformità delle pavimentazioni aeroportuali nel tempo. Per ottimizzare le risorse disponibili, risulta infatti necessario identificare le sezioni della pavimentazione che necessitano di manutenzione, la soluzione manutentiva maggiormente appropriata, il momento ideale per l'intervento ed il corrispondente budget richiesto. Al fine di identificare le sezioni che necessitano di riabilitazione, i dati strutturali delle pavimentazioni sono solitamente ottenuti mediante prove distruttive su campioni ottenuti tramite carotaggi e scavi sull'esistente [Chern et al., 2005]. Queste procedure sono economicamente molto onerose in quanto comportano la successiva riparazione del foro di prova e possono causare la chiusura dell'area di indagine, anche per diversi giorni [Huang et al., 2004]. Fortunatamente, il progresso tecnologico sia lato hardware che software ha reso possibile sostituire tali procedure con tecnologie più rapide e meno invasive, che non compromettono l'integrità strutturale della pavimentazione. Questi nuovi approcci, noti anche come Test Non Distruttivi (*Non Destructive Tests* - NDT), consentono ugualmente di determinare le proprietà fisiche e meccaniche di una pista di volo aeroportuale, ma in modo non distruttivo [Hoffman and Thompson, 1982]. Per questo motivo, negli ultimi anni c'è stato un ampio utilizzo di tali tecnologie nei settori della progettazione e relativa diagnosi delle pavimentazioni sia stradali che aeroportuali [Bush III and Alexander, 1985, Sebaaly et al., 1986, Chang et al., 1992, Weil et al., 1992, Ullidtz and Coetzee, 1995]. Tra i metodi NDT, quello più comunemente utilizzato per la valutazione delle condizioni delle infrastrutture aeroportuali è il deflettometro dinamico pesante a massa battente (*Heavy Weight Deflectometer* - HWD) [Gopalakrishnan, 2012].

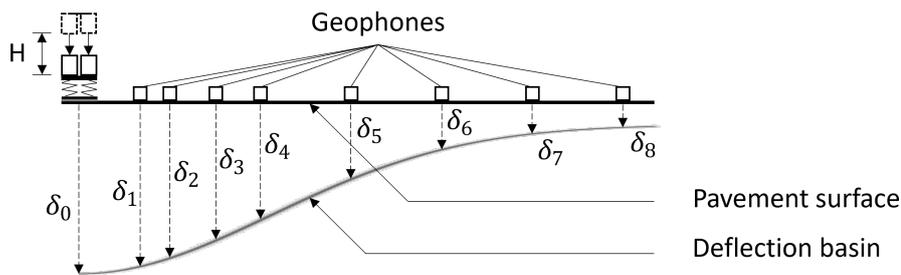


Figura 6.1: Schematizzazione di una prova tramite HWD.

L'HWD simula i carichi elevati applicati dalle ruote degli aeromobili [USDOT, 2011] e consente di effettuare rapidamente test sull'intera pavimentazione aeroportuale in modo che possano essere definiti i conseguenti programmi di manutenzione e riabilitazione. Successivamente all'applicazione del carico, la curva di deflessione risultante viene misurata mediante geofoni

posti a molteplici distanze dall'asse di carico ed è rappresentativa della risposta complessiva del sistema, oltre che un ottimo indicatore del comportamento meccanico della pavimentazione. È pratica comune elaborare successivamente le deflessioni misurate (Figura 6.1 al fine di ottenerne una significativa combinazione, i parametri del bacino di deflessione (*Deflection Basin Parameters* - DBPs) [Talvik and Aavik, 2009] che caratterizzano ancor meglio l'eventuale degrado strutturale della pista di volo aeroportuale [Gopalakrishnan and Thompson, 2007].

Secondo la letteratura scientifica di riferimento, il metodo più tradizionale per elaborare le deflessioni da HWD è la *back-calculation*, o calcolo a ritroso, per la stima della capacità portante della pavimentazione. Conosciuto anche come problema di identificazione dei parametri, il calcolo a ritroso è essenzialmente un processo di ottimizzazione eseguito per ottenere una mappatura inversa di una relazione costitutiva nota utilizzando punti noti discreti o continui [Goktepe et al., 2006]. La *back-calculation* consiste quindi in un'analisi numerica delle deflessioni misurate al fine di stimare la capacità portante delle pavimentazioni aeroportuali investigate. Per raggiungere questo obiettivo, le deflessioni misurate vengono confrontate con quelle calcolate. Queste ultime vengono determinate sulla base della teoria del multistrato elastico (*Multi-Layer Elastic* - MLE), in cui le tensioni e le deformazioni sono caratterizzate da equazioni differenziali del quarto ordine [Burmister, 1945]. Nella pratica, a partire dallo spessore degli strati, dall'intensità del carico e da alcuni moduli di rigidità sintetici (ovvero assunti pari a valori basati sull'esperienza e sulle migliori pratiche ingegneristiche), è possibile calcolare le conseguenti deflessioni. Il processo di adattamento viene eseguito in modo iterativo fino a quando si ottiene la migliore corrispondenza tra le deflessioni calcolate e quelle misurate [Ullidtz, 2000, Lytton, 1989, Uzan, 1994]. A causa delle semplificazioni introdotte dalla teoria MLE, alcuni studi hanno osservato che i risultati prodotti da tale calcolo a ritroso sono accettabili solo per lo strato superficiale della pavimentazione, mentre i moduli di rigidità di fondazione e sottofondo sono spesso sottovalutati o sovrastimati [Siddharthan et al., 1991]. Ciò si verifica come conseguenza del fatto che i software tradizionali di *back-calculation* trascurano gli effetti dinamici dei carichi prodotti dall'HWD, che, per essere implementati, richiederebbero una elevata complessità del modello e dei notevoli oneri computazionali [Goel and Das, 2008]. Per questo motivo, diverse tecniche di apprendimento automatico (su tutte, le reti neurali artificiali) sono state sempre più spesso utilizzate per risolvere problemi di questo tipo [Scimemi et al., 2016]. Queste astrazioni funzionali della struttura neurale biologica sono particolarmente adatte ad affrontare problemi difficili da comprendere senza una teoria formale ben definita per la loro soluzione [Adeli, 2001]. Anche se le ANNs lavorano sulla base di un approccio non fisicamente basato, sono caratterizzate da un'elevata efficienza computazionale e da errori di previsione particolarmente ridotti [Li and Wang, 2019]. È importante sottolineare che l'elaborazione neurale dei dati è molto rapida e consente di eseguire le analisi delle deflessioni da HWD in modo efficiente addirittura in situ [Gopalakrishnan et al., 2009]. Tuttavia, le reti neurali artificiali richiedono solitamente un ampio set di dati per essere adeguatamente addestrate. Per questo motivo, a partire da simulazioni agli elementi finiti, vengono spesso preparati dei database sintetici che verranno successivamente utilizzati come input dei modelli neurali al fine di consentire loro di addestrarsi opportunamente [Wang et al., 2021].

L'obiettivo del presente studio è quello di sviluppare un modello predittivo in grado di migliorare la *back-calculation* convenzionale mediante avanzate funzionalità di apprendimento

automatico. Il risultato sarà un innovativo strumento di elaborazione dati che, utilizzando tecniche all'avanguardia di *data augmentation* e l'efficacia computazionale delle reti neurali artificiali, è in grado prevedere con elevata affidabilità valori di modulo di rigidezza del conglomerato bituminoso utilizzato per lo strato superficiale della pista di volo (E_{AC}), anche in punti non ancora campionati. Poiché tramite l'approccio proposto sarà possibile prevedere il modulo di rigidezza fornendo in input le deflessioni e la distribuzione spaziale dei punti di campionamento, la relazione funzionale tra le variabili in gioco è stata rispettata. Rispetto all'uso tradizionale delle reti neurali artificiali nel campo dell'ingegneria delle pavimentazioni, l'approccio proposto evita le preliminari e lunghe simulazioni agli elementi finiti solitamente impiegate per addestrare il modello. Pertanto, tale approccio rapido ma ugualmente preciso potrebbe essere utilizzato in un APMS come strumento di supporto per la pianificazione delle priorità di intervento. Fornendo una stima numerica del modulo di rigidezza in una posizione arbitraria della pista di volo, il modello predittivo proposto potrebbe consentire di identificare le aree che richiedono con più urgenza degli interventi di manutenzione, riducendo così i costi del monitoraggio strumentale e consentendo una gestione attiva ed efficiente dell'infrastruttura [Justo-Silva et al., 2021].

6.2 Campagna Sperimentale

6.2.1 Indagini in situ

La campagna sperimentale oggetto di studio è stata condotta presso l'aeroporto "Falcone e Borsellino" di Palermo-Punta Raisi, in un periodo compreso tra il 9 ed il 29 febbraio 2012. L'aeroporto internazionale italiano si trova 35 km a ovest della provincia di Palermo. L'infrastruttura è costituita da due piste che si intersecano: la principale denominata Pista 07/25 e quella secondaria denominata Pista 02/20. Su quest'ultima, sono state raccolte misure di deflessione tramite un deflettometro dinamico pesante a massa battente.

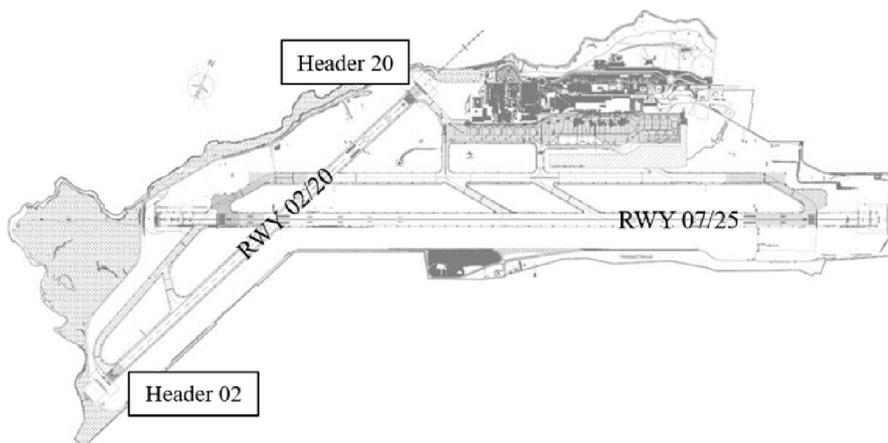


Figura 6.2: Disposizione delle piste di volo dell'Aeroporto di Palermo.

Sono state individuate cinque direttrici di rilevamento: l'asse centrale della pista (0m) e altri assi distanti $\pm 3\text{m}$ e $\pm 6\text{m}$ da esso. La porzione di pista considerata, che parte dalla testata 02 a sud e arriva alla testata 20 a nord, era lunga 1800 m [Figura 6.2]. Gli esperimenti sono stati eseguiti a intervalli regolari di 100 m, ottenendo così 19 punti di impatto per ciascun asse longitudinale di misurazione. Il numero complessivo di punti di rilevamento è risultato dunque pari a 95 [Figura 6.3].

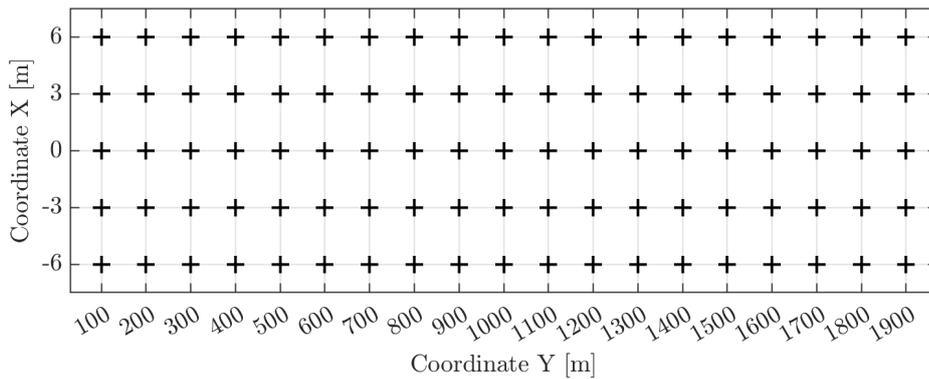


Figura 6.3: Punti di impatto individuati sulla pista di volo 02/20.

La strumentazione ad alto rendimento impiegata è rappresentata dal Dynatest 8000, un dispositivo in grado di eseguire test non distruttivi ed ampiamente utilizzato per determinare le proprietà fisico-meccaniche delle pavimentazioni aeroportuali. Il principio è quello di indurre piccole deflessioni superficiali della pavimentazione applicando un carico controllato di tipo impulsivo. Ciò al fine di simulare fedelmente gli effetti che la ruota di un aereomobile esercita sull'infrastruttura sottostante [Crovetti et al., 1989]. Il carico è ottenuto facendo cadere una massa in sospensione da un'altezza predeterminata su una piastra di 30 cm di diametro appoggiata sulla superficie della pavimentazione. L'entità del carico impulsivo trasmesso dal dispositivo alla pavimentazione può variare da 30 kN a 240 kN modificando sia il peso che l'altezza della caduta. Nel caso in esame, è stata adottata un'entità dell'impulso pari a circa 140 kN. Insieme alla piastra, una serie di trasduttori accelerometrici è posta a contatto con la pavimentazione ed è in grado di misurare le deflessioni indotte a diverse distanze dall'asse di carico, quali: 0 mm (δ_0), 200 mm (δ_1), 300 mm (δ_2), 450 mm (δ_3), 600 mm (δ_4), 900 mm (δ_5), 1200 mm (δ_6), 1500 mm (δ_7) e 1800 mm (δ_8) dal centro della piastra. La strumentazione HWD si ritiene di consueto preferibile rispetto alle altre poiché è particolarmente semplice, affidabile ed economica [Claessen et al., 1976]. Infatti, l'utilizzo di test non distruttivi evita i costi che sarebbero associati alla ricostruzione dell'area di intervento se si adottassero prove distruttive [Chen et al., 2003]. Una preliminare mappatura dei valori δ_0 misurati al di sotto della piastra di carico [Figura 6.4] mostra immediatamente che le deflessioni più elevate si registrano in prossimità dell'asse centrale della pavimentazione. Ciò consente di individuare le cosiddette *Touchdown Zones* (TDZ), ossia quelle porzioni di infrastruttura in cui gli aerei in atterraggio sono destinati a toccare la pista in un primo momento [ICAO, 2016]. In queste zone, i valori di δ_0 superano i 900 μm e sono più che doppi o talvolta tripli rispetto alle deflessioni misurate

allontanandosi verso le estremità della pista. Questa considerazione fornisce una valutazione preliminare circa quali aree della pista potrebbero necessitare dei più urgenti interventi di manutenzione.

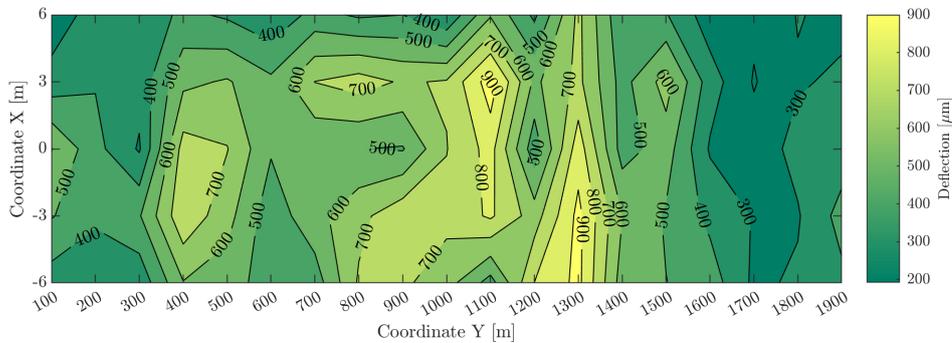


Figura 6.4: Mappatura delle deflessioni δ_0 registrate nei punti di impatto.

6.2.2 Parametri del bacino di deflessione - DBPs

I parametri del bacino di deflessione si ottengono elaborando i risultati prodotti da un'indagine HWD e talvolta possono essere utilizzati per monitorare l'integrità strutturale di una pavimentazione aeroportuale in esercizio [Hossain and Zaniewski, 1991]. Sulla base di un'estesa revisione della letteratura, i più ampiamente utilizzati sono:

- Indice di curvatura della superficie (*Surface Curvature Index* - SCI), fornisce un'indicazione sulle variazioni di resistenza degli strati prossimi alla superficie.

$$SCI_1 = \delta_0 - \delta_2 \quad (6.2.1)$$

$$SCI_2 = \delta_0 - \delta_3 \quad (6.2.2)$$

$$SCI_3 = \delta_0 - \delta_4 \quad (6.2.3)$$

Il valore di SCI è un ottimo indicatore delle condizioni dello strato superficiale della pavimentazione e, per determinati spessori, i valori di modulo E_{AC} e quelli di SCI tendono a mostrare una relazione approssimativamente lineare in una scala bilogarithmica [Xu et al., 2002].

- Rapporto di deflessione (*Deflection Ratio* - DR), tiene conto del tipo e della qualità dei materiali correlandoli attraverso il rapporto di due deflessioni.

$$DR = \frac{\delta_4}{\delta_0} \quad (6.2.4)$$

- Area sottesa dal bacino di deflessione (*Area Under Deflection Basin Curve* - AREA), mette in relazione la rigidezza della pavimentazione con un fattore di forma ed è pari all'area

sottesa dalla curva del bacino di deflessione normalizzata rispetto a δ_0 usando la regola di Simpson [Gopalakrishnan and Thompson, 2005]. Tale formulazione risulta valida quando le deflessioni sono espresse utilizzando il sistema internazionale di unità di misura.

$$AREA = \frac{150(\delta_0 + 2\delta_2 + 2\delta_4 + \delta_5)}{\delta_0} \quad (6.2.5)$$

- Area sotto il profilo della pavimentazione (*Area Under Pavement Profile - AUPP*) [Hill, 1988] un ulteriore indicatore sensibile alle proprietà dello strato superficiale della pavimentazione, come riscontrato da Thompson e Garg [Thompson and Garg, 1997].

$$AUPP = \frac{5\delta_0 + 2\delta_2 + 2\delta_4 + \delta_5}{\delta_0} \quad (6.2.6)$$

6.2.3 Procedura di back-calculation

Per determinare i moduli di rigidezza dello strato superficiale della pavimentazione nei punti di impatto è stato utilizzato il metodo RO.M.E. (*ROad Moduli Evaluation*). Questa procedura fa riferimento alla teoria del multistrato elastico, all'equazione di Boussinesq-Odemark e al metodo dello spessore equivalente. Infatti, a partire dalle misure di spessore degli strati che compongono la pista di volo (ottenute da carotaggi e radargrammi [Figura 6.5]) e dai risultati HWD ottenuti durante la campagna sperimentale, il metodo RO.M.E. consente di determinare lo stato di sollecitazione/deformazione di ciascun punto dello strato bituminoso assunto come omogeneo, isotropo e di spessore semi-infinito. Successivamente, attraverso una procedura iterativa, il metodo RO.M.E. rende il bacino di deflessione teorico congruente con quello misurato sperimentalmente, in modo da poter stimare i moduli di rigidezza degli strati di pavimentazione. Per una descrizione completa ed esaustiva delle operazioni del metodo RO.M.E. si rimanda al contributo di Battiato et al. [Battiato et al., 1994].

I moduli ottenuti tramite *back-calculation* [Figura 6.6] mostrano chiaramente che le aree caratterizzate dalle deflessioni più elevate presentano dei valori di modulo di rigidezza più bassi. Allo stesso modo, deflessioni più contenute sono correlate a valori di modulo di rigidezza più alti. Inoltre, la determinazione preventiva degli spessori dei diversi strati della pavimentazione ha permesso di individuare le cosiddette sezioni omogenee, ovverosia quelle sezioni della pista caratterizzate dalla stessa stratigrafia. Una variabile categorica che tenga conto delle sezioni omogenee sarà di seguito denominata HS (*Homogeneous Section*). I moduli calcolati a ritroso e la variabile HS sono stati quindi associati, per ogni punto di impatto, alle corrispondenti deflessioni ed ai rispettivi DBPs. Questi dati costituiranno parte della tabella Input/Target necessaria per implementare la successiva strategia di apprendimento supervisionato.

6.3 Modellazione predittiva con approcci neurali superficiali

In questo specifico contesto, si è deciso di analizzare le performance di tre potenziali modelli neurali caratterizzati ciascuno da un'architettura di tipo $4 - n - 1$, $8 - n - 1$ e $9 - n - 1$. Il primo numero corrisponde ai neuroni che costituiscono lo strato di input (un neurone per

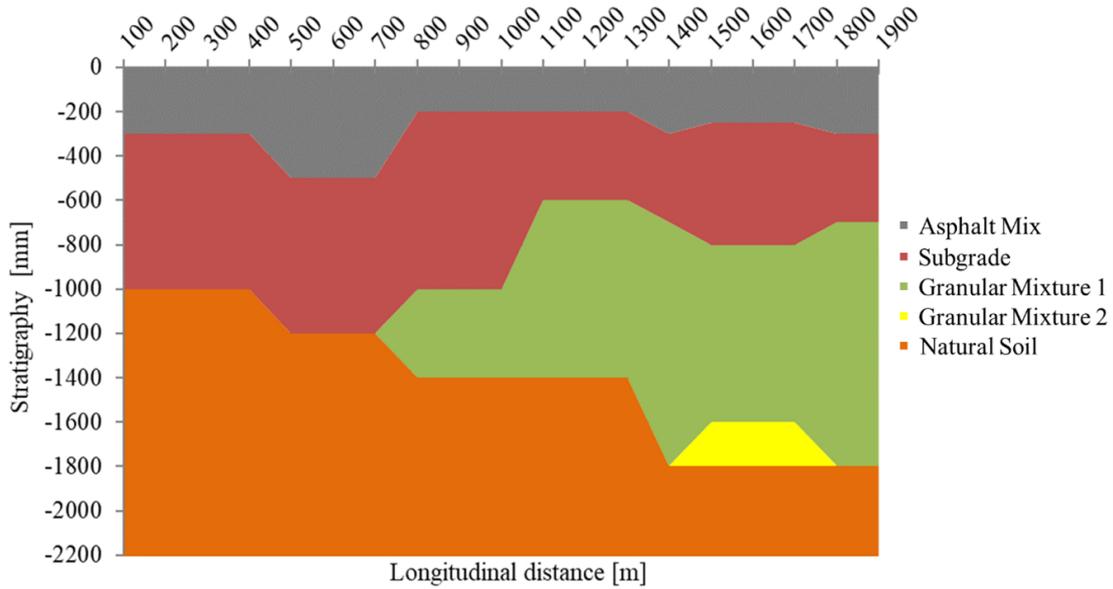


Figura 6.5: Spessore degli strati di pavimentazione ottenuti mediante carotaggi.

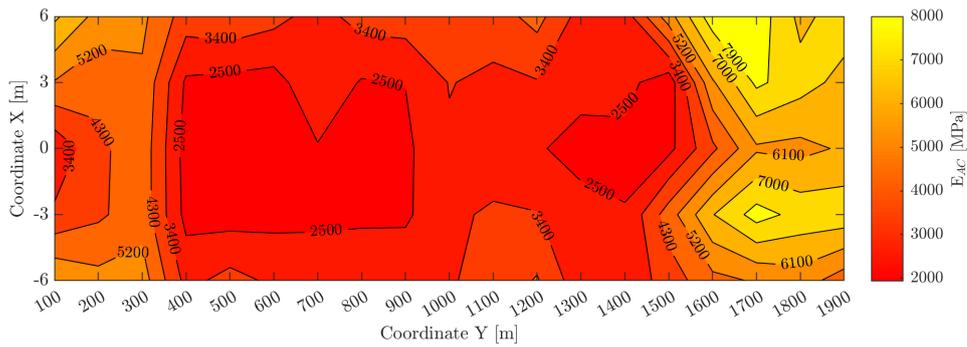


Figura 6.6: Mappatura dei moduli di rigidità dello strato superficiale calcolati a ritroso.

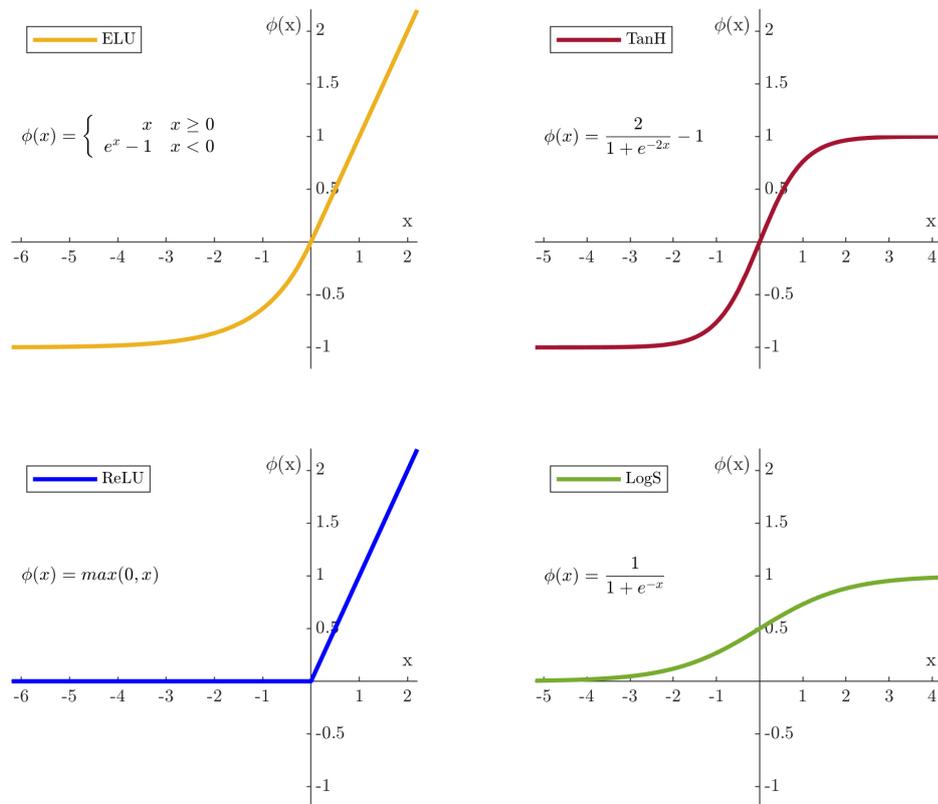


Figura 6.7: Funzioni di attivazione ELU, ReLU, TanH e LogS.

ciascuna variabile in ingresso), mentre n indica il numero di neuroni dello strato nascosto con n un numero intero compreso tra 1 e 30. Una volta addestrato e testato il primo modello (a partire da $n = 1$), sono stati iterativamente generati modelli con $n + 1$ neuroni nello strato nascosto fino ad ottenere una struttura composta da 30 neuroni nascosti. La migliore funzione di attivazione per ciascun modello è stata investigata all'interno di un gruppo di quattro funzioni distinte: esponenziale lineare (ELU), lineare rettificata (ReLU), tangente iperbolica (TanH) e sigmoide logistica (LogS). Le rispettive espressioni analitiche, insieme ai grafici risultanti, sono mostrate nella Figura 6.7. Ne è risultata una procedura di ricerca a griglia volta ad individuare la migliore combinazione di neuroni nascosti-funzione di attivazione che garantisca le migliori prestazioni della rete neurale all'interno degli intervalli di ricerca. Infine, lo strato di uscita è sempre stato composto da un unico neurone ed associato ad una funzione di attivazione lineare.

Ciascun modello presentava in input la variabile categorica identificativa delle sezioni omogenee HS e le coordinate spaziali dei punti di impatto sulla pista (X,Y). Inoltre, il primo modello neurale aggiungeva al vettore delle variabili di input la corrispondente misura di deflessione δ_0 . Il secondo modello considerava sia δ_0 che tutte le altre misure di deflessione

comprese tra δ_2 e δ_5 . Il terzo ed ultimo modello prendeva in considerazione i valori dei DBPs definiti al paragrafo 6.2.2 piuttosto che i valori delle singole deflessioni. La singola variabile di output predetta consisteva nel modulo di rigidità dello strato superficiale della pavimentazione aeroportuale (E_{AC}). Ancor prima di essere assegnati alla rete, sia gli input che gli output sono stati sottoposti ad una procedura di standardizzazione. Ciò significa che, per ciascuna variabile, ogni osservazione è stata sottratta dal rispettivo valore di media e diviso per la corrispondente deviazione standard. Tale operazione è stata eseguita poiché l'algoritmo scelto per addestrare la rete (Regolarizzazione Bayesiana) è più efficace quando gli input ed i target sono scalati in modo da ricadere approssimativamente nell'intervallo $[-1,1]$ [Beale et al., 2010].

6.3.1 Regolarizzazione Bayesiana

Il processo di addestramento di una rete neurale consiste nell'applicare le operazioni sequenziali necessarie per regolare con precisione i pesi sinaptici e le distorsioni (bias), al fine di generalizzare le risposte prodotte dai suoi output. [Da Silva et al., 2017]. Nella presente ricerca, si è partiti dalla tabella Input/Target menzionata nel paragrafo 6.2.3, e per questo motivo l'intera procedura viene definita supervisionata. Questa prevede due fasi fondamentali: una "in avanti" ed una "all'indietro". La prima consiste nell'assegnare alla rete il vettore delle caratteristiche (di input) \mathbf{x} (ad esempio, l'insieme delle osservazioni registrate) e nel calcolare l'output corrispondente $\hat{\mathbf{y}}$. La seconda, invece, consiste nel confrontare l'output generato con il target desiderato \mathbf{y} . La differenza tra $\hat{\mathbf{y}}$ e \mathbf{y} definirà la cosiddetta funzione di perdita $F(\hat{\mathbf{y}}, \mathbf{y})$ che sarà successivamente impiegata per determinare le correzioni da applicare alla matrice dei pesi e dei bias \mathbf{W} . Esistono diverse espressioni analitiche che definiscono le modalità di aggiornamento di \mathbf{W} in base al valore assunto dalla funzione di perdita per un prefissato numero di iterazioni E , differenziando così i diversi algoritmi di apprendimento. La funzione di perdita F più comunemente impiegata è l'MSE. La sua espressione analitica è presentata nell'equazione 3.8.5. Il gradiente della funzione di perdita F rispetto a \mathbf{W} calcolato applicando un algoritmo di retro-propagazione (*backpropagation* [Rumelhart et al., 1986]), permette di aggiornare i pesi della rete in modo da minimizzare il valore di perdita. Osservando l'equazione 6.3.1, si può comprendere come i pesi della iterazione successiva ($e + 1$) vengano calcolati a partire da quelli utilizzati nell'iterazione precedente $\forall e \in \{1, \dots, E\}$:

$$\mathbf{W}^{e+1} = \mathbf{W}^e - (\nabla^2 F(\mathbf{W}^e))^{-1} \nabla F(\mathbf{W}^e) \quad (6.3.1)$$

Combinando questa equazione con l'algoritmo di retro-propagazione Levenberg-Marquardt (LM) [Hagan and Menhaj, 1994], si ottiene:

$$\mathbf{W}^{e+1} = \mathbf{W}^e - [\mathbf{J}^T(\mathbf{W}^e)\mathbf{J}(\mathbf{W}^e) + \mu_e\mathbf{I}]^{-1} \mathbf{J}^T(\mathbf{W}^e)\mathbf{v}(\mathbf{W}^e) \quad (6.3.2)$$

dove \mathbf{W} rappresenta la matrice dei pesi e dei bias, \mathbf{J} rappresenta la matrice Jacobiana della funzione di perdita F rispetto a \mathbf{W}^e , \mathbf{I} rappresenta la matrice identità e \mathbf{v} rappresenta il vettore degli errori della rete calcolati attraverso l'espressione:

$$\mathbf{v}(\mathbf{W}^e) = \hat{\mathbf{y}}(\mathbf{W}^e) - \mathbf{y} \quad (6.3.3)$$

Lo scalare μ (noto con il nome di tasso di apprendimento) definisce il tasso di convergenza dell'algoritmo. All'aumentare di μ , l'algoritmo LM si sposta di un piccolo passo nella direzione di discesa più ripida $\mathbf{J}^T(\mathbf{W}^e)\mathbf{v}(\mathbf{W}^e)$. Al contrario, al diminuire di μ , la velocità di convergenza dell'algoritmo aumenterebbe con la possibilità che però l'algoritmo possa oltrepassare il minimo. Per questo motivo, durante la fase di addestramento, il valore di μ viene progressivamente modificato in modo da raggiungere la convergenza il più rapidamente possibile, evitando minimi locali indesiderati. È necessario impostare un valore iniziale di μ per la prima iterazione. Successivamente, questo aumenta (o diminuisce) essendo moltiplicato per un fattore $\mu_{inc} > 1$ (o $\mu_{dec} < 1$) nel caso in cui l'iterazione precedente abbia condotto a peggiori (o migliori) risultati in termini di funzione di perdita F . In questo modo, iterazione dopo iterazione, la funzione di perdita tende a diminuire gradualmente. Infine, è necessario impostare un valore massimo di μ (denominato μ_{max}) in modo da interrompere l'addestramento se tale valore viene superato. Quando viene raggiunto μ_{max} (oppure, in alternativa, al termine delle iterazioni E) i valori migliori di pesi e bias vengono identificati e mantenuti fissi. Successivamente, il vettore degli input destinato alla fase di test viene assegnato alla rete, facendola lavorare solo in avanti e determinando così l'indice di perdita del modello su dati mai elaborati prima. Tuttavia, in questo modo, è possibile incorrere in un fenomeno noto nel *Machine Learning* come *overfitting* [Padhi et al., 2019]. Ciò si verifica quando i valori dei pesi sinaptici sono troppo elevati e, per questo motivo, si è deciso di implementare una tecnica di regolarizzazione che tenga conto dei valori della matrice \mathbf{W} . Tale tecnica consiste nel riscrivere la funzione di perdita non più come semplice MSE ma come segue:

$$F_{opt}(\hat{\mathbf{y}}(\mathbf{W}^e), \mathbf{y}, \mathbf{W}^e) = \beta \|\hat{\mathbf{y}}(\mathbf{W}^e) - \mathbf{y}\|_2^2 + \alpha \|\mathbf{W}^e\|_2^2 \quad (6.3.4)$$

L'operatore $\|\cdot\|_2^2$ rappresenta la norma 2 ed è applicata prima agli errori $\mathbf{v}(\mathbf{W}^e) = \hat{\mathbf{y}}(\mathbf{W}^e) - \mathbf{y}$ e poi ai parametri della rete \mathbf{W}^e . Il primo termine è anche noto come somma dei quadrati degli errori (*Sum Squared Error* - SSE) mentre il secondo termine rappresenta una correzione che tiene conto della complessità della rete attraverso i suoi pesi, ed è noto come penalità. Inoltre, entrambi i termini sono moltiplicati da parametri indicati come β e α , il cui rapporto determina la "regolarità" (*smoothness*) della funzione di perdita. Maggiore è il rapporto α/β , più "regolare" sarà la risposta della rete. Tale concetto fu introdotto per la prima volta da Tikhonov [Tikhonov, 1963]. Il termine di penalità (o regolarizzazione) costringe la funzione di perdita risultante ad essere regolare. Quando i pesi sono grandi, la funzione di perdita può avere grandi pendenze ed è quindi più probabile che si adatti eccessivamente ai dati di addestramento. D'altra parte, se i pesi sono piccoli, la funzione di perdita interpolerà in modo omogeneo i dati di addestramento, ottenendo buone prestazioni anche al di fuori del set di addestramento. Una scelta corretta del rapporto di regolarizzazione α/β è imprescindibile per produrre una rete in grado di generalizzare in maniera corretta. In questo caso, per definire correttamente i termini di penalità, si è deciso di utilizzare l'approccio di David MacKay [MacKay, 1992]. Tale approccio consiste in una prima inizializzazione dei parametri α e β (oltre che dei pesi della rete). Mediante tecniche statistiche e di ottimizzazione Bayesiana, i valori di α and β vengono variati ad ogni iterazione, modificando così la funzione di perdita e, ovviamente, anche il suo punto di minimo. Generalmente, ogni qualvolta si stabilisce un nuovo minimo, i parametri di regolarizzazione risultano sempre più accurati. Alla fine, la precisione sarà abbastanza elevata da far sì che

la funzione obiettivo non cambi più in modo significativo ad ogni iterazione successiva e la convergenza sarà quindi stata raggiunta. Il set di iperparametri utilizzato per la prima iterazione è stato ottenuto utilizzando i valori standard per gli iperparametri implementati all'interno dell'algoritmo LM previsto da MATLAB®Toolbox. Nello specifico, i valori di μ , μ_{inc} , μ_{dec} , μ_{max} e E sono stati posti rispettivamente pari a 0.001, 10, 0.1, 10^{10} e 1000. Come descritto in precedenza, la migliore combinazione di dimensione dello strato nascosto e della corrispondente funzione di attivazione è stata identificata tramite una ricerca a griglia. Infine, un ultimo parametro w_w è stato definito. Esso rappresenta il numero di ri-addestramenti eseguiti per ciascuna iterazione ed è stato impostato pari a 10. Questo al fine di ottenere la rete migliore tra le 10 reti configurate come risultato di un partizionamento k-fold per ogni combinazione di neuroni nascosti-funzione di attivazione. Il valore di k è stato quindi impostato pari a 5 per essere in linea con la letteratura scientifica di riferimento [Kuhn et al., 2013, Li and Wang, 2019]. Pertanto, l'80% dei dati disponibili è stato utilizzato per la fase di training set mentre il rimanente 20% è stato impiegato per quella di test.

6.3.2 Data Augmentation

Nel contesto dell'apprendimento automatico, si parla di *Data Augmentation* ogni qualvolta si utilizzi una particolare tecnica per ampliare le dimensioni del set di dati di partenza. Si tratta di una generazione di dati sintetici che proviene dall'elaborazione delle informazioni raccolte e non da una nuova campagna sperimentale. In questo senso, incrementare la quantità di dati a disposizione della rete aumenta le sue capacità predittive, migliorando la sua generalizzazione. Le tecniche di aumento dei dati sono particolarmente diffuse nel campo della classificazione delle immagini, poiché ruotare, zoomare, ritagliare o modificare la luminosità non significa alterare le informazioni contenute all'interno di un'immagine. Allo stesso modo, nel caso della *data augmentation* per le serie temporali è necessario utilizzare tecniche che non disturbino le informazioni raccolte durante la campagna sperimentale ma che, allo stesso tempo, permettano di aumentare la dimensione del campione. Allo stesso modo, nel caso della *data augmentation* per le serie temporali è necessario utilizzare tecniche che non disturbino le informazioni raccolte durante la campagna sperimentale ma che, allo stesso tempo, permettano di aumentare la dimensione del campione. Poiché la raccolta di nuovi dati può risultare particolarmente costosa e dispendiosa in termini di tempo, le tecniche di interpolazione multipla si stanno progressivamente affermando nell'ambito della *data augmentation*. Con il termine interpolazione si intende un procedimento che consenta di stimare valori sconosciuti utilizzando dati noti. Più precisamente, se la funzione $f(x)$ della variabile reale x è sconosciuta mentre sono noti i valori $f(x_i)$ che tale funzione assume in due o più punti $x_i (i = 1, 2, \dots, n)$ all'interno di un predeterminato intervallo, la stima di un valore della funzione $f(x)$ per qualsiasi punto x intermedio prende il nome di interpolazione [Lunardi, 2018]. Conformemente a quanto suggerito da Oh et al. [Oh et al., 2020], è opportuno che la dimensione del set di dati noti non sia mai inferiore a quella del set di dati ottenuti tramite interpolazione. Per questo motivo, si è deciso di pressoché raddoppiare la dimensione del set di dati registrati interpolando la funzione bidimensionale sconosciuta in un punto intermedio (croci rosse) tra due punti di impatto successivi appartenenti alla stessa linea di misura [Figura 6.8].

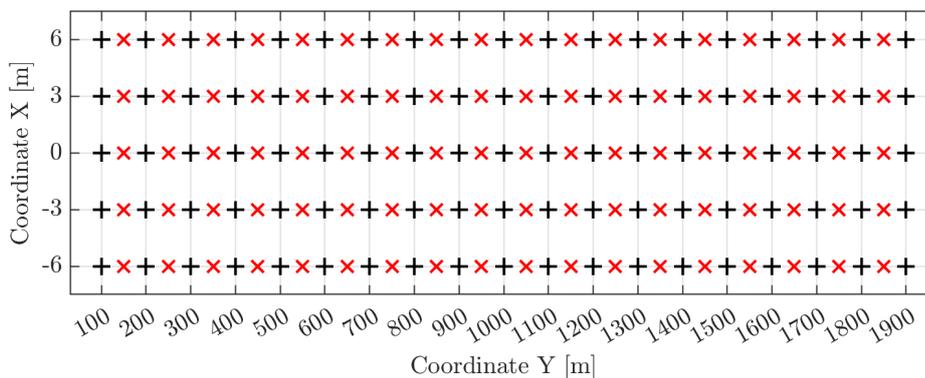


Figura 6.8: Punti di impatto sintetici ottenuti tramite Data Augmentation.

La funzione scelta come interpolatore diventa una sorta di iperparametro del modello. Il metodo qui esposto utilizza l'interpolazione Akima modificata, spesso nota come tecnica "makima". In generale, l'algoritmo di Akima esegue un'interpolazione cubica per produrre polinomi continui a tratti con derivate continue del primo ordine [Akima, 1970]. A differenza dell'interpolazione polinomiale di 3° grado o della spline cubica, l'algoritmo di Akima evita oscillazioni locali eccessive, pur essendo in grado di gestire dati a carattere oscillatorio. L'algoritmo Akima modificato (cioè "makima") rappresenta un'estensione del metodo appena menzionato ed è stato progettato per interpolare tramite una funzione bivariata regolare $z = z(x, y)$ valori registrati in punti di una griglia rettangolare piana. La funzione interpolante è quindi un polinomio bicubico in ogni cella della griglia rettangolare [Akima, 1974]. Lo scopo principale resta quello di produrre meno oscillazioni possibili all'interno dei punti della griglia. Le osservazioni ottenute attraverso l'interpolazione makima (85 dati sintetici) sono state utilizzate esclusivamente durante la fase di addestramento del modello e sono state aggiunte al partizionamento k-fold per un totale di 161 dati di addestramento. Assumendo come riferimento un modello neurale messo a punto tramite MATLAB®Toolbox, è stato effettuato un confronto tra tale modello semplificato e quello proposto nell'ambito di questa ricerca. Tale confronto ha consentito di valutare il miglioramento della stima dei valori di rigidità fornita dalla rete neurale artificiale proposta. Inoltre, assumendo che l'implementazione della *data augmentation* sia quasi equivalente ad una campagna sperimentale più fitta, si può intuire come la variazione della risoluzione delle misure HWD influisca sul risultato finale del modello.

6.4 Discussione dei risultati

Sebbene il dataset di partenza presentasse un'elevata variabilità lungo la pista sia in termini di deflessioni misurate che di moduli di rigidità calcolati a ritroso (a causa delle pronunciate differenze di comportamento meccanico), il modello neurale proposto ha fornito risultati molto soddisfacenti. Le sue prestazioni sono state valutate in termini di coefficiente di correlazione di Pearson R, di errore quadratico medio MSE e di coefficiente di determinazione rettificato R_{adj}^2 .

Come noto, la prima metrica esprime - se esiste - una correlazione lineare tra i moduli calcolati a ritroso E_{AC} e quelli predetti dal modello neurale \hat{E}_{AC} :

$$R(E_{AC}, \hat{E}_{AC}) = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{E_{AC_i} - \mu_{E_{AC}}}{\sigma_{E_{AC}}} \right) \left(\frac{\hat{E}_{AC_i} - \mu_{\hat{E}_{AC}}}{\sigma_{\hat{E}_{AC}}} \right) \quad (6.4.1)$$

con μ e σ a rappresentare rispettivamente la media e la deviazione standard delle variabili corrispondenti. Valori di R superiori a 0.80 sono considerati tipici di una correlazione significativa [Benesty et al., 2009]. La seconda metrica fornisce un'accurata stima delle capacità di generalizzazione del modello attraverso il calcolo della media dei quadrati delle differenze tra moduli di rigidezza calcolati a ritroso e quelli previsti dal modello neurale:

$$MSE(E_{AC}, \hat{E}_{AC}) = \frac{1}{n} \sum_{i=1}^n (E_{AC_i} - \hat{E}_{AC_i})^2 \quad (6.4.2)$$

La terza metrica, invece, indica in quale percentuale le variabili di input sono in grado di giustificare una variazione della variabile di output:

$$R_{adj}^2(E_{AC}, \hat{E}_{AC}) = 1 - \left(\frac{n-1}{n-p} \right) \frac{SSE}{SST} \quad (6.4.3)$$

con $SSE = \sum_{i=1}^n (E_{AC_i} - \hat{E}_{AC_i})^2$ a rappresentare la somma dei quadrati degli errori, $SST = \sum_{i=1}^n (E_{AC_i} - \mu_{E_{AC}})^2$ la devianza totale di E_{AC} , n il numero complessivo di osservazioni, e p il numero di input del modello. La sua formulazione permette di valutare se l'aggiunta di più variabili indipendenti in input si traduca in un miglioramento della bontà di adattamento del modello di regressione. Infatti, la metrica R_{adj}^2 aumenta qualora i nuovi termini migliorino la bontà di adattamento del modello, mentre diminuisce qualora questo miglioramento non sia apprezzabile. Per quanto riguarda l'approccio di ricerca a griglia, tra tutte le possibili combinazioni di neuroni nascosti-funzione di attivazione, è stata selezionata quella che massimizzasse l' R_{adj}^2 . L'intento è stato quello di produrre una rete che commettesse piccoli errori e, allo stesso tempo, utilizzasse al meglio i parametri forniti in input. Da questo punto di vista, il modello che ha prodotto i risultati migliori è stato il terzo. L'architettura 9-23-1 con funzione di attivazione logistico-sigmoide (di seguito denominata LogS-SNN) ha prodotto un valore di R_{adj}^2 pari a 0,9516 [Tabella 6.1], confermando che i parametri del bacino di deflessione, insieme alle sezioni omogenee e alle coordinate spaziali dei punti di impatto, siano dati di partenza particolarmente adatti per addestrare un modello predittivo in termini di modulo di rigidezza.

Per rendere più apprezzabile il processo di apprendimento del miglior modello neurale, in Figura 6.9 viene illustrato l'andamento grafico delle metriche R ed MSE . Si può osservare come la prima metrica cresca progressivamente fino all'ottimo per poi assestarsi su valori costanti. Allo stesso modo, la seconda metrica diminuisce fino all'ottimo per poi stabilizzarsi su valori costanti. Tali considerazioni rendono chiaro come aggiungere ulteriori neuroni allo strato nascosto oltre la soglia del valore ottimale appesantirebbe computazionalmente il modello senza apportare alcun beneficio significativo in termini di performance predittive.

Un'ulteriore dimostrazione dell'efficienza della rete dal punto di vista dei costi computazionali è stata fornita dal valore del numero effettivo di parametri (*Effective Number of Parameters* -

Tabella 6.1: Sintesi delle performance predittive dei migliori modelli neurali sviluppati.

Inputs	Output	Funz. attiv.	Architettura	R	MSE	R_{adj}^2
δ_0, HS, X, Y	E_{AC}	ELU	4-12-1	0.9673	0.0682	0.9231
		ReLU	4-27-1	0.9408	0.1309	0.8627
		TanH	4-25-1	0.9780	0.0494	0.9477
		LogS	4-22-1	0.9772	0.0474	0.9461
$\delta_0, \delta_2, \delta_3, \delta_4, \delta_5,$ HS, X, Y	E_{AC}	ELU	8-27-1	0.9805	0.0423	0.9368
		ReLU	8-3-1	0.9455	0.1217	0.8277
		TanH	8-9-1	0.9806	0.0437	0.9370
		LogS	8-13-1	0.9844	0.0370	0.9493
$SCI_1, SCI_2, SCI_3, DR,$ AUPP, AREA, HS, X, Y	E_{AC}	ELU	9-18-1	0.9804	0.0501	0.9303
		ReLU	9-14-1	0.9555	0.0963	0.8441
		TanH	9-26-1	0.9807	0.0439	0.9312
		LogS	9-23-1	0.9864	0.0321	0.9516

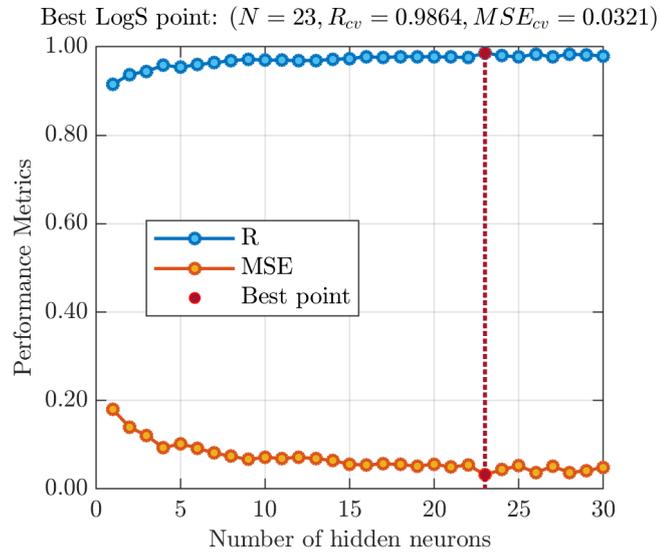


Figura 6.9: Metriche di performance del modello LogS-SNN.

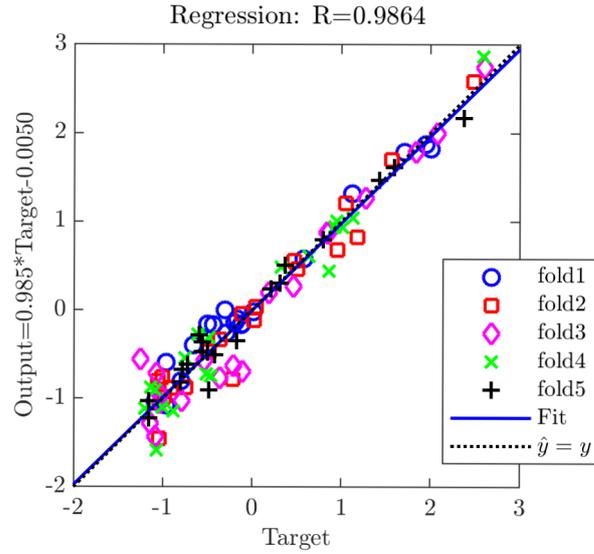


Figura 6.10: Rappresentazione delle performance del modello LogS-SNN per ciascun fold.

ENP). Tale metrica fornisce un'indicazione circa quanti parametri (pesi e bias) del modello neurale sono effettivamente utilizzati per ridurre la funzione di perdita [Hagan and Menhaj, 1994]. ENP è espresso da:

$$ENP = \sum_{i=1}^5 \frac{\gamma_i - 2\alpha_i \text{tr}(\nabla^2 F(\mathbf{W}_i))}{\gamma_i} \quad (6.4.4)$$

con \mathbf{W}_i e γ_i che rappresentano rispettivamente la matrice di pesi e bias ed il numero totale di parametri dell' i -esimo modello SNN addestrato durante la procedura di cross-validation. La traccia della matrice Hessiana può essere calcolata a partire dalla matrice Jacobiana degli errori commessi in fase di addestramento e dei parametri α_i e β_i , come spiegato in [Hagan and Menhaj, 1994]. Con un valore pari a 185/254, il valore di ENP ottenuto per il modello LogS-SNN ha dimostrato come, nonostante l'elevato numero di neuroni nello strato nascosto, più del 70% dei parametri del modello venga effettivamente utilizzato per ridurre la funzione di perdita. Si tratta probabilmente della combinazione ottimale di efficienza computazionale e accuratezza nelle previsioni, come evidenziato anche dal valore dell'errore quadratico medio MSE_{CV} pari a 0,0321. Per una completa rappresentazione dei risultati, le prestazioni del modello LogS-SNN per ciascuno dei 5 fold sono state rappresentate graficamente [Figura 6.10]. Come spiegato nel paragrafo 3.6.2, calcolando la media dei risultati prodotti per ciascuno dei 5 fold, è possibile valutare le effettive capacità predittive del modello, ottenendo un R_{CV} pari a 0,9864.

Utilizzando un VivoBookProN580GD-FI018T con CPU Intel(R) core(TM) i7-8750H @2,20GHz e 16GB di RAM, questo modello impiega circa 13 secondi per elaborare i dati e permette di prevedere il valore di modulo di rigidità \hat{E}_{AC} in un qualsiasi punto della pista. Similmente a quanto fatto al paragrafo 6.2.3, è possibile dunque generare una mappatura dei moduli di

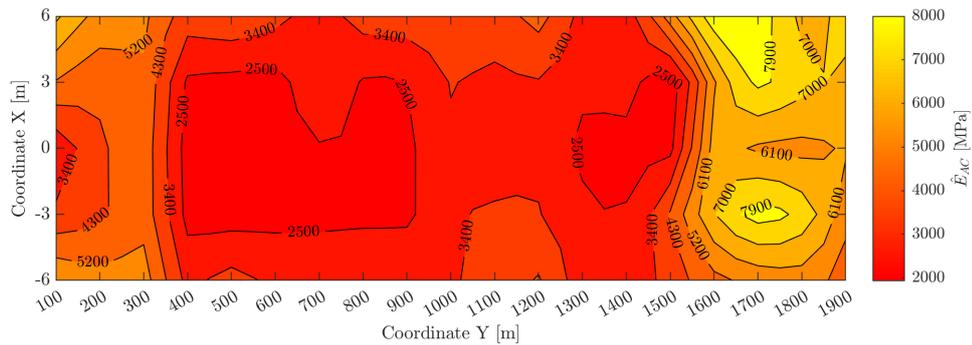


Figura 6.11: Mappatura dei moduli di rigidezza predetti dal modello LogS-SNN.

rigidezza dello strato superficiale della pista di volo, questa volta predetti dal modello neurale, fornendo una stima quantitativa delle prestazioni meccaniche della pavimentazione al momento delle indagini sperimentali HWD [Figura 6.11]. È opportuno sottolineare che una semplice interpolazione dei dati di modulo calcolati a ritroso può certamente stimare il valore di E_{AC} in un punto qualsiasi della pista, ma costituirebbe una predizione di capacità portante effettuata senza considerare alcuna relazione fenomenologica. Il modello neurale proposto, invece, fornisce previsioni del modulo di rigidezza in qualsiasi punto della pavimentazione, ma richiede in input le deflessioni misurate (o delle loro elaborazioni matematiche), le loro coordinate lungo la pista e la corrispondente stratigrafia. In questo senso, almeno da un punto di vista logico, l'approccio proposto conserva la relazione fondamentale tra la capacità portante e le variabili menzionate.

6.5 Conclusioni

Il lavoro di modellazione predittiva descritto in questo capitolo è stato incentrato sulla previsione del modulo di rigidezza dello strato superficiale di una pista di volo. In questo senso, la capacità portante dell'infrastruttura aeroportuale rappresenta un elemento critico nella gestione della pavimentazione stessa e dell'intero *terminal*. L'approccio metodologico sviluppato è stato basato su algoritmi di apprendimento automatico e su tecniche di *data augmentation*. In particolare, a scopi modellistici, sono state implementate reti neurali artificiali poco profonde (SNN) caratterizzate da un'architettura a tre strati ed un algoritmo di interpolazione spaziale noto come makima (o Akima modificato). I risultati di una campagna sperimentale condotta mediante HWD sulla pista di volo 02/20 dell'Aeroporto di Palermo hanno permesso di sviluppare tre potenziali modelli neurali predittivi. Le coordinate spaziali dei punti di impatto ed una variabile categorica identificativa del tipo di sezione omogenea della pavimentazione sono stati sempre inclusi all'interno del vettore di input di ciascun modello. In aggiunta a tali variabili, il primo modello SNN ha considerato anche la misura di deflessione δ_0 . Nel secondo modello SNN sono state incluse anche tutte le deflessioni comprese tra δ_2 e δ_5 . Nel terzo ed ultimo modello SNN, le misure di deflessione sono state sostituite dai parametri del bacino di deflessione. Per ottimizzare il numero di neuroni nascosti e selezionare la funzione di attivazione più adatta agli scopi proposti, è stata effettuata un'estesa ricerca a griglia. Infine, per identificare

il modello caratterizzato dalle migliori performance predittive, sono stati implementati algoritmi di regolarizzazione Bayesiana ed una procedura di k-fold cross-validation.

I risultati hanno evidenziato come il terzo modello SNN abbia dimostrato di possedere le migliori capacità predittive: questo era caratterizzato da 23 neuroni nello strato nascosto e da una funzione di attivazione sigmoide logistica. Le previsioni generate in termini di modulo di rigidità sono state molto accurate, caratterizzate da valori di coefficiente di correlazione di Pearson R e MSE rispettivamente pari a 0.9864 e 0.0321. Inoltre, al fine di considerare contemporaneamente sia la dimensione del dataset disponibile che il numero di parametri del modello, è stata introdotta la metrica R_{adj}^2 . Il suo valore è risultato pari a 0.9516, confermando pienamente le buone prestazioni del modello predittivo sviluppato.

Il modello neurale messo a punto tramite MATLAB®Toolbox ha prodotto un coefficiente di correlazione di Pearson pari a 0.9460, decisamente inferiore al valore ottenuto in seguito all'applicazione delle tecniche di *data augmentation*. Il coefficiente R è infatti risultato pari a 0.9864, con un significativo aumento del 4%. L'approccio di apprendimento automatico proposto si è dimostrato in grado di prevedere il modulo di rigidità dello strato superficiale della pista di volo in un qualsiasi punto della stessa. Pertanto, può diventare un elemento cruciale nei cosiddetti sistemi di gestione delle pavimentazioni aeroportuali, permettendo così di ottimizzare il budget disponibile da destinare alle operazioni di manutenzione. In questo modo è possibile migliorare sia la sicurezza che la sostenibilità delle infrastrutture aeroportuali. L'approccio proposto è stato sviluppato in relazione alla pista di volo di Palermo: tuttavia, potrebbe essere facilmente adottato per analizzare qualsiasi pista di volo o area pavimentata (come anche parcheggi), a patto che siano disponibili misure deflettometriche. Ad oggi, lo studio non ha considerato serie storiche di deflessioni. Sarebbe dunque interessante implementare queste informazioni aggiuntive al fine di valutare se la metodologia qui presentata sia adatta non solo a valutare lo stato di deterioramento attuale della pavimentazione, ma se sia anche in grado di prevederne l'evoluzione per programmare nel tempo le priorità di intervento.

Capitolo 7

Caratterizzazione Prestazionale di Conglomerati Bituminosi

” *Data are becoming the new raw material of business.*

– Craig Mundie

7.1 Introduzione

Le diverse tipologie di conglomerati bituminosi (denominati anche *bituminous mixtures* - BM) utilizzati per le pavimentazioni devono essere adeguatamente progettati come miscele di aggregati e bitume in modo da resistere ai carichi del traffico ed alle diverse condizioni atmosferiche. Caratteristiche meccaniche e proprietà volumetriche inadeguate, infatti, possono condurre a molteplici tipologie di danni nelle pavimentazioni che generalmente comprendono fessure dovute alla fatica o alle basse temperature, deformazioni permanenti, *stripping* ecc. Tali forme di dissesto riducono la vita utile della pavimentazione e rappresentano un serio problema per la sicurezza degli utenti della strada. Di conseguenza, è fondamentale caratterizzare correttamente il comportamento meccanico dei conglomerati bituminosi in relazione alla loro composizione, al fine di consentire, durante la fase di progettazione della miscela, un'ottimizzazione della stessa basata sulle prestazioni (*performance-based mix design*) [Zhou et al., 2004, Gandomi et al., 2011, Alavi et al., 2011]. Attualmente, per valutare le prestazioni dei conglomerati bituminosi, si utilizzano metodi sperimentali che richiedono costose prove di laboratorio e tecnici competenti altamente specializzati [Dias et al., 2014, Liu and Wu, 2014, Garcia et al., 2014, Pasandín and Pérez, 2015, Zaumanis et al., 2016, Wang et al., 2017]. Di conseguenza, qualsiasi variazione della composizione delle miscele in esame, in termini di tipo o contenuto di bitume, piuttosto che di gradazione degli aggregati, richiede che vengano effettuate nuove prove di laboratorio con un conseguente aumento dei tempi e dei costi del processo di progettazione. Recentemente, molti ricercatori hanno dedicato il loro impegno al problema della definizione di un modello matematico o numerico del comportamento meccanico dei conglomerati bituminosi, in grado di produrre rapidamente una previsione affidabile della risposta della miscela in esame. Al fine di sviluppare equazioni predittive, si

possono utilizzare due principali tipologie di procedura, ovvero avanzati metodi di modellazione costitutiva piuttosto che approcci statistici o di *data science*. Sebbene i metodi costitutivi meccanici permettano di effettuare un'analisi razionale ed approfondita della risposta del materiale [Erkens et al., 2002, Giunta and Angela Pisano, 2006, Underwood and Kim, 2011, Yun and Kim, 2013, Pasetto and Baldo, 2015, Di Benedetto et al., 2016, Pasetto and Baldo, 2016, Darabi et al., 2019], gli approcci statistici e quelli di apprendimento automatico, sebbene non siano fisicamente basati, stanno riscuotendo un notevole successo all'interno della comunità accademica grazie alla loro considerevole attendibilità previsionale [Kim and Kim, 2006, Laurinavičius and Oginskas, 2006, Shukla and Das, 2008, Rahman et al., 2019]. Tuttavia, nel corso degli anni, si è osservato che le regressioni statistiche dei dati sperimentali producano spesso previsioni meno accurate rispetto ai metodi di apprendimento automatico, in particolar modo rispetto alle reti neurali artificiali [Specht et al., 2007, Mirzahosseini et al., 2011, Androjić and Marović, 2017, Alrashyda and Abo-Qudais, 2018, Montoya and Haddock, 2019]. In linea generale, come discusso all'interno del paragrafo 4.3, le ANN simulano modelli semplificati del cervello umano e la loro potenza di calcolo deriva dal numero di connessioni stabilite tra le unità di calcolo fondamentali, ovvero i neuroni artificiali. La loro funzione principale è quella di identificare le relazioni tra input e output di un campione sperimentale rappresentativo, imitando il processo di apprendimento biologico. Tali modelli neurali si basano fondamentalmente su un approccio di adattamento non lineare, né fisico né meccanicistico, al fine di correlare i dati sperimentali. Negli ultimi anni, un numero crescente di ricercatori ha utilizzato le ANN in molteplici applicazioni nell'ambito dell'ingegneria civile, ottenendo risultati sorprendenti anche in termini di valutazione delle proprietà e delle prestazioni delle pavimentazioni in conglomerato bituminoso. A titolo puramente illustrativo si riportano i seguenti esempi: nel 2005, [Tarefder et al., 2005] hanno sviluppato una rete neurale *feedforward* profonda (caratterizzata da quattro strati nascosti) per correlare i parametri di progettazione della miscela con le prestazioni dei campioni di conglomerato bituminoso in termini di permeabilità; nel 2008, [Ozsahin and Oruc, 2008] hanno messo a punto modelli neurali per determinare la relazione esistente tra il modulo resiliente dell'emulsione bituminosa ed i fattori che lo influenzano, come ad esempio il tempo di maturazione, la quantità di cemento ed il contenuto di bitume residuo; nel 2009, [Tapkın et al., 2009] hanno presentato un'applicazione neurale per la previsione dei risultati di prove di *creep* su conglomerati bituminosi modificati con polipropilene; previsioni accurate sono state inoltre prodotte in termini di vita a fatica dei conglomerati bituminosi sottoposti a differenti condizioni climatiche e di carico [Xiao et al., 2009, Ahmed et al., 2017]; infine, [Ceylan et al., 2009] hanno discusso l'accuratezza e la robustezza dei modelli basati su reti neurali artificiali per la stima del modulo dinamico dei conglomerati bituminosi miscelati a caldo. Sebbene gli studi menzionati abbiano tentato di introdurre nuovi approcci alla progettazione empirico-meccanica delle miscele, l'approccio Marshall è ancora ampiamente adottato in molti laboratori di tecnica stradale [Aksoy et al., 2012, Van Thanh and Feng, 2013, Abdoli et al., 2015, Sarkar et al., 2016, Xu et al., 2016, Zumrawi and Khalill, 2017, Al-ammari et al., 2019]. Di conseguenza, [Tapkın et al., 2010] hanno verificato la possibilità di applicare le reti neurali artificiali per la previsione dei risultati delle prove Marshall condotte sui conglomerati bituminosi. Il modello ANN da loro proposto utilizza le proprietà fisiche dei campioni Marshall per prevederne i valori di stabilità, scorrimento e di quoziente Marshall ottenuti al termine delle prove

meccaniche. In linea generale, le caratteristiche meccaniche delle miscele bituminose dipendono dalle proprietà volumetriche e dal contenuto di bitume. Tali parametri devono chiaramente rispettare i limiti stabiliti dalle specifiche normative vigenti ma, in ogni caso, i vuoti nell'aggregato minerale, i vuoti riempiti di bitume ed i vuoti d'aria (AV) sono determinati tramite specifici test (normati secondo [UNI EN 12697-8, 2019]) che richiedono tempi e costi aggiuntivi. In questo senso, [Khuntia et al., 2014] hanno proposto un modello di rete neurale che utilizza le quantità di bitume e di aggregato contenute all'interno dei provini Marshall per prevedere i valori di stabilità e scorrimento Marshall, oltre che dei vuoti d'aria ottenuti dalle prove. Allo stesso modo, [Zavrtanik et al., 2016] hanno utilizzato le reti neurali artificiali per stimare il contenuto di vuoti d'aria in diverse tipologie di conglomerato bituminoso prodotte secondo normativa [UNI EN 13108-1, 2016]. Nonostante le diverse ANN abbiano fornito con successo equazioni predittive per accelerare la progettazione empirica Marshall delle miscele, tali modelli computazionali si basano solitamente su una struttura di rete neurale impostata a priori dall'ingegnere-ricercatore ed addestrata su un sottoinsieme casuale della banca dati disponibile. Nel caso di un insieme di dati relativamente piccolo, tale pratica può comportare il rischio di tralasciare alcuni *trend* rilevanti dall'insieme di addestramento e provocare un errore previsionale variabile misurato sull'insieme di osservazioni di test, a causa della variabilità del campione di dati e dell'architettura selezionata per il modello neurale [James et al., 2013]. Questi problemi possono però essere evitati se si effettua un'efficiente selezione del modello ed un'appropriata ripartizione dei dati. In particolare, la ricerca dell'architettura di rete ottimale, uno dei compiti più difficili negli studi che coinvolgono le ANN, consiste nell'identificare gli iperparametri che producono le migliori prestazioni sull'insieme di osservazioni di test. All'interno della letteratura scientifica di riferimento [Mirzahosseini et al., 2011, Androjić and Marović, 2017], è possibile trovare applicazioni circa le procedure *trial-and-error*, come le ricerche casuali o a griglia, per identificare gli iperparametri ottimali di un algoritmo di apprendimento automatico per un dato problema di modellazione predittiva. Tuttavia, un approccio talvolta maggiormente efficiente prende il nome di ottimizzazione bayesiana (BO) ed è stato particolarmente apprezzato negli ultimi anni [Shahriari et al., 2015]. Dato che la determinazione del valore della funzione di performance al variare dei diversi iperparametri rappresenta una procedura estremamente onerosa, l'approccio Bayesiano costruisce un modello probabilistico, chiamato "surrogato", che mappa gli iperparametri delle precedenti valutazioni con la probabilità di ottenere un determinato punteggio sulla funzione di performance ed utilizza tale modello per identificare la serie successiva di iperparametri promettenti (che ottimizzano la funzione surrogata) da valutare sull'effettiva funzione di performance [Bergstra et al., 2011, Bergstra et al., 2013].

Il presente capitolo si propone di sviluppare una procedura autonoma ed imparziale di selezione del modello neurale per problemi di modellazione predittiva del comportamento meccanico delle miscele bituminose, utilizzando un metodo BO che sostituisca la più onerosa procedura *trial-and-error*. In particolare, un approccio neurale è stato utilizzato per analizzare la rigidità, la stabilità Marshall, lo scorrimento Marshall ed il contenuto di vuoti d'aria di 320 campioni Marshall testati presso il Laboratorio di Ingegneria Autostradale dell'Università Aristotele di Salonicco. La banca dati sperimentale comprende diversi tipi di bitume e di aggregati, includendo un'ampia gamma di contenuti di bitume e granulometrie di aggregati. Inoltre, sono stati impiegati conglomerati preparati sia in laboratorio che in impianto ed il loro

sito di produzione è stato considerato tra le variabili di input del modello neurale proposto; esso correla le proprietà meccaniche e volumetriche (rilevate mediante prove di laboratorio) alle caratteristiche fondamentali delle miscele bituminose, come ad esempio il contenuto di bitume (% in peso della miscela), il rapporto tra filler e bitume (%), i tipi di bitume e di aggregato e la massima granulometria nominale.

7.2 Campagna Sperimentale

7.2.1 Materiali

320 provini Marshall, con un diametro di 100 mm e un'altezza di 63,7 mm, sono stati prodotti sia in laboratorio che in impianto secondo quanto previsto dalla normativa [UNI EN 12697-30, 2018]. Tali miscele, progettate nell'ambito di un progetto di ricerca condotto presso l'Università Aristotele di Salonicco, sono state prodotte utilizzando molteplici contenuti di bitume e diverse granulometrie di aggregati. Gli aggregati erano di tipo calcareo o diabase, caratterizzati da dimensioni nominali massime di 20 mm o 12,5 mm. Per monitorarne le proprietà fisiche, sono state condotte diverse prove sperimentali i cui risultati sono riportati in Tabella 7.1.

Tabella 7.1: Proprietà degli aggregati.

Proprietà	Normativa	Tipo di aggregato	
		Calcare	Diabase
Coeff. Los Angeles (%)	EN 1097-2	29	25
Valore di levigabilità (%)	EN 1097-8	-	55 ÷ 60
Indice di appiattimento (%)	EN 933-3	23	18
Equivalente in sabbia (%)	EN 933-8	79	59
Blu di metilene (mg/g)	EN 933-9	3.3	8.3

I conglomerati bituminosi con dimensione nominale massima dell'aggregato pari a 12,5 mm (BM12.5) sono stati progettati per strati di *binder* mentre quelli con dimensione nominale massima pari a 20 mm (BM20) sono stati progettati per strati di base. La banca dati qui riportata contiene osservazioni relative a 27 granulometrie differenti che soddisfano i limiti stabiliti dalle attuali normative greche rispettivamente per strati di *binder* e di base. Per quanto riguarda la produzione delle miscele in laboratorio, sono state impiegate 4 diverse granulometrie per soddisfare i requisiti delle BM12.5 ed altre 4 per soddisfare i requisiti delle BM20. Le rimanenti granulometrie riguardano miscele BM20 preparate in impianto. Le Figure 7.1, 7.2 e 7.3, mostrano una rappresentazione grafica delle curve granulometriche considerate. Un bitume standard di penetrazione 50/70 è stato impiegato nella preparazione di 129 campioni Marshall, mentre i restanti 191 sono stati prodotti utilizzando un bitume modificato in laboratorio tramite aggiunta di polimeri SBS (*Styrene-Butadiene-Styrene*). Entrambe le tipologie di bitume sono state testate per garantire che le proprietà fisiche fossero conformi agli specifici requisiti di accettazione. I risultati ottenuti sono stati riportati in Tabella 7.2.

Tabella 7.2: Proprietà dei bitumi.

Proprietà	Normativa	Tipo di bitume	
		50/70	Modificato
Penetrazione (0.1·mm)	EN 1426	64	45
Punto di rammollimento (°C)	EN 1427	45.6	78.8
Recupero elastico (%)	EN 13398	-	97.5
Punto di rottura Fraass (°C)	EN 12593	- 7.0	- 15.0

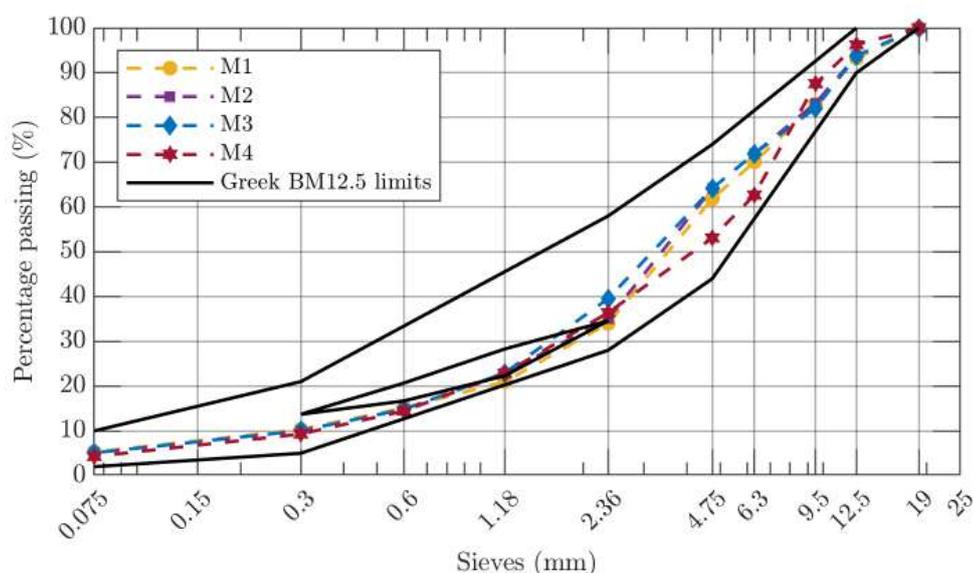


Figura 7.1: Curve granulometriche delle miscele BM12.5 preparate in laboratorio.

7.2.2 Indagini di laboratorio

I campioni Marshall sono stati prodotti utilizzando una percentuale di bitume compresa tra il 3.8 % ed il 6.0% (in peso della miscela), ed in un numero pari a tre per ciascun contenuto di bitume adottato. La Tabella 7.3 riassume il numero di campioni prodotti per ciascuna combinazione di bitume ed aggregato; inoltre, viene riportata un'abbreviazione per l'identificazione di ciascuna miscela preparata. Tra tutti i parametri meccanici relativi ai conglomerati bituminosi, l'ITSM (*Indirect Tensile Stiffness Modulus*) consente di effettuare una caratterizzazione razionale e prestazionale delle miscele in esame [Dias et al., 2014, Xiao and Amirkhanian, 2009]. Pertanto, il test ITSM (Figura 7.4) è stato eseguito su tutti i campioni utilizzando le condizioni standard di prova: temperatura pari a 20 °C, deformazione target fissata a 5 μm e tempo di crescita pari a 124 ms, così come definite dalla normativa [UNI EN 12697-26, 2018]. Successivamente, tenuto conto che i parametri Marshall sono ancora ampiamente utilizzati nella progettazione delle pavimentazioni stradali [Aksoy et al., 2012, Van Thanh and Feng, 2013, Abdoli et al., 2015, Sarkar et al., 2016, Xu et al., 2016, Zumrawi and Khalill, 2017, Al-ammari et al., 2019, Tapkın et al., 2010], sono sta-

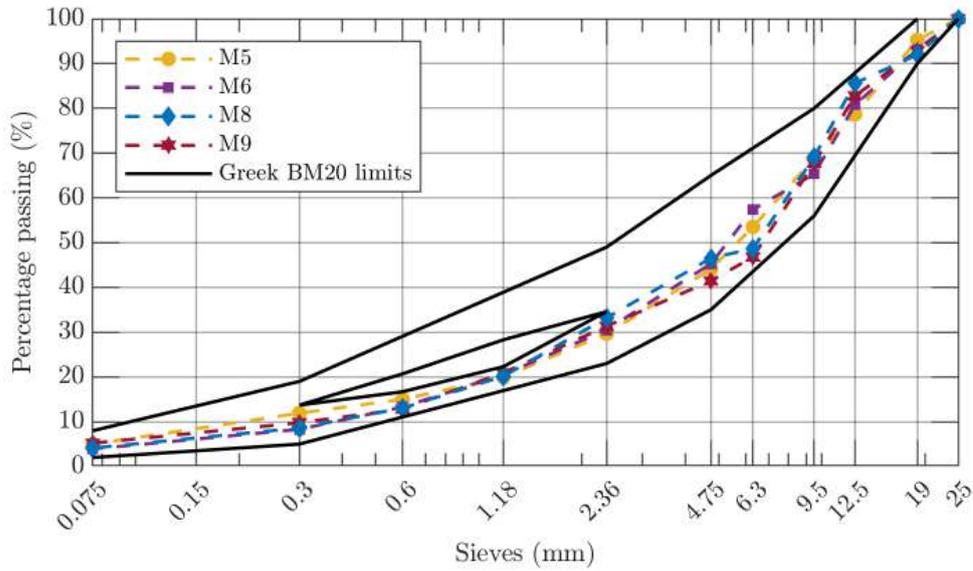


Figura 7.2: Curve granulometriche delle miscele BM20 preparate in laboratorio.

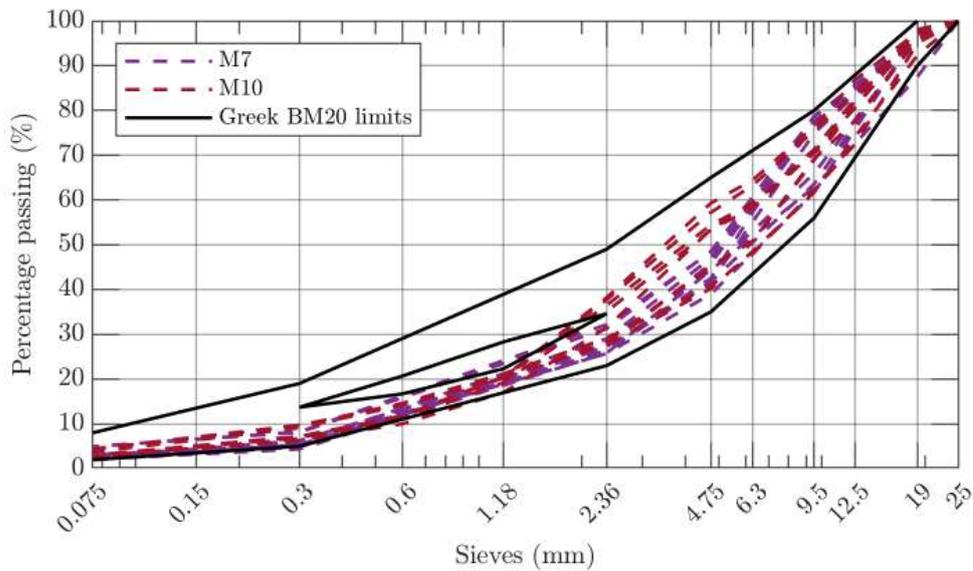


Figura 7.3: Curve granulometriche delle miscele BM20 preparate in impianto.

ti valutati la stabilità (*Marshall Stability* - MS) e lo scorrimento (*Marshall Flow* - MF) Marshall per le miscele bituminose prodotte, secondo la normativa [UNI EN 12697-34, 2020]. Infine, le caratteristiche volumetriche dei provini sono state determinate secondo normativa [UNI EN 12697-8, 2019]. In Tabella 7.4, è possibile osservare alcune informazioni statistiche

CAPITOLO 7. CARATTERIZZAZIONE PRESTAZIONALE DI CONGLOMERATI BITUMINOSI

(valori minimi, massimi, medi e di deviazione standard) riferite alla caratteristiche meccaniche ed alle proprietà volumetriche dei conglomerati bituminosi investigati.

Tabella 7.3: Sintesi dei campioni Marshall preparati.

Dimensione nominale massima (mm)	Tipo di aggregato	Tipo di bitume	Luogo di produzione	ID della miscela	Campioni
12.5	Calcare	50/70	Laboratorio	M1	30
12.5	Calcare	Modificato	Laboratorio	M2	30
12.5	Diabase	50/70	Laboratorio	M3	30
12.5	Diabase	Modificato	Laboratorio	M4	30
20	Calcare	50/70	Laboratorio	M5	39
20	Calcare	Modificato	Laboratorio	M6	30
20	Calcare	Modificato	Impianto	M7	41
20	Diabase	50/70	Laboratorio	M8	30
20	Diabase	Modificato	Laboratorio	M9	30
20	Diabase	Modificato	Impianto	M10	30



Figura 7.4: Strumentazione di laboratorio per test ITSM.

Tabella 7.4: Caratteristiche meccaniche e volumetriche dei conglomerati investigati.

ID	Parametro	Minimo	Massimo	Media	Dev. Std
M1	ITSM (MPa)	3756	5554	4556.43	567.93
	MS (kN)	7.71	12.17	9.93	1.09
	MF (mm)	1.99	4.70	3.18	0.84
	AV (%)	1.77	6.37	3.99	1.37
M2	ITSM (MPa)	3628	5142	4345.50	486.50
	MS (kN)	8.74	14.00	11.35	1.73
	MF (mm)	2.00	4.20	3.20	0.57
	AV (%)	2.20	6.29	4.18	1.23
M3	ITSM (MPa)	3812	5942	4804.10	725.03
	MS (kN)	10.30	15.20	12.88	1.53
	MF (mm)	2.00	5.00	3.35	0.95
	AV (%)	1.49	8.91	5.22	2.38
M4	ITSM (MPa)	4035	6293	5076.17	759.06
	MS (kN)	11.60	16.43	13.62	1.42
	MF (mm)	2.20	5.00	3.40	0.92
	AV (%)	1.33	8.36	5.05	2.18
M5	ITSM (MPa)	3215	4919	4252.26	502.24
	MS (kN)	8.91	14.86	11.37	1.51
	MF (mm)	2.18	4.60	3.15	0.50
	AV (%)	2.17	6.75	4.28	1.16
M6	ITSM (MPa)	3907	6043	5243.10	538.97
	MS (kN)	10.40	13.99	11.81	1.21
	MF (mm)	2.24	4.16	3.24	0.40
	AV (%)	1.68	5.21	3.49	1.08
M7	ITSM (MPa)	3103	6399	5065.34	906.93
	MS (kN)	6.60	14.75	9.86	2.20
	MF (mm)	2.10	9.86	3.22	0.62
	AV (%)	3.03	2.20	5.22	1.19
M8	ITSM (MPa)	2304	4900	3829.63	783.23
	MS (kN)	10.45	15.48	13.05	1.36
	MF (mm)	2.20	5.00	3.37	0.83
	AV (%)	0.35	8.44	4.37	2.43
M9	ITSM (MPa)	2930	5994	4911.30	851.15
	MS (kN)	8.92	15.48	12.22	2.03
	MF (mm)	1.85	5.00	3.10	0.72
	AV (%)	1.26	8.44	4.97	1.86
M10	ITSM (MPa)	4094	5968	5309.27	565.88
	MS (kN)	7.80	16.55	11.98	2.15
	MF (mm)	2.70	5.40	3.95	0.76
	AV (%)	4.60	9.70	7.12	1.59

7.3 Modellazione predittiva con approcci neurali profondi

In questo specifico contesto, il modello di rete neurale proposto si compone di 3 neuroni nello strato di input a rappresentare rispettivamente il contenuto di bitume (*Bitumen content* - BC), il rapporto filler su bitume (*Filler-to-bitumen ratio* - FtB), ed una variabile categorica identificativa del conglomerato bituminoso in termini di tipo di bitume, dimensione nominale massima dell'aggregato e luogo di produzione (valori pari a 1 per M1, 2 per M2 e così via). Le caratteristiche descritte rappresentano parametri fondamentali dell'ingegneria delle pavimentazioni in quanto strettamente legati alla composizione delle miscele. Ogni strato nascosto del modello neurale è stato equipaggiato con N neuroni artificiali ed una funzione di attivazione $\phi(\cdot)$ a scelta tra ELU, ReLU e TanH. Infine, lo strato nascosto, è stato definito mediante l'utilizzo di 4 neuroni artificiali (corrispondenti rispettivamente ai parametri ITSM, MS, MF e AV) e di una funzione di attivazione identità. Studi precedenti hanno già permesso di effettuare previsioni di diversi parametri Marshall; tuttavia, ciascun parametro risultava predetto in maniera distinta mediante l'utilizzo di uno specifico modello [Xiao and Amirkhanian, 2009, Tapkın et al., 2010, Ozgan, 2011, Khuntia et al., 2014, Zavrtnik et al., 2016, Baldo et al., 2018]. Invece, nel contesto appena delineato, le previsioni avvengono in maniera simultanea tramite l'utilizzo di un modello neurale definito *multi-output*.

7.3.1 Regolarizzazione L2

Con l'obiettivo di aumentare la velocità di convergenza dell'algoritmo di addestramento del modello messo a punto, è stato implementato l'algoritmo di ottimizzazione dei pesi sinaptici per reti neurali profonde noto come *Ranger optimizer*. Tale algoritmo consente di combinare i benefici di due metodologie migliorative note in letteratura come *RAdam optimizer* [Liu et al., 2019] e rettifica con approccio *Lookahead* [Zhang et al., 2019], qui solo menzionate per ragioni legate alla semplificazione della trattazione. Inoltre, per evitare la presenza di spiacevoli fenomeni di *overfitting*, è stata implementata la tecnica nota come Regolarizzazione L2 [Ng, 2004]. In tal senso, la funzione di perdita (L) è stata modificata ed ottimizzata come segue:

$$L_{opt}(\hat{y}^{(d)}, y^{(d)}, W^{(e-1)}) = L(\hat{y}^{(d)}, y^{(d)}) + \beta \|W^{(e-1)}\|_2^2 \quad (7.3.1)$$

con $\hat{y}^{(d)}$ e $y^{(d)}$ a rappresentare rispettivamente le previsioni effettuate dal modello ed i valori obiettivo, $W^{(e-1)}$ la matrice dei pesi relativi all'iterazione $e - 1$ e β a rappresentare un iperparametro utile a regolare l'entità del termine $\|W^{(e-1)}\|_2^2$.

Anche in questo caso, la procedura di convalida incrociata *k-fold* è stata adottata, utilizzando - come suggerito da letteratura - un valore di k pari a 5.

7.3.2 Ottimizzazione Bayesiana degli iperparametri

Nell'ambito del problema in esame, ovvero la previsione simultanea delle proprietà meccaniche e volumetriche dei conglomerati bituminosi per pavimentazioni stradali, è possibile definire la funzione obiettivo $f(\cdot)$ come media del coefficiente di correlazione di Pearson, tale che $f : X_L \times X_N \times X_{act} \times X_\alpha \times X_\beta \times X_E \rightarrow [-1, 1]$. Pertanto, lo scopo ultimo sarà massimizzare

$f(\cdot)$ in funzione dei sei diversi iperparametri appena definiti: L , numero di strati nascosti, N , numero di neuroni artificiali all'interno di ciascuno strato nascosto, act , funzione di attivazione di ciascuno strato nascosto, α , tasso di apprendimento, β iperparametro di regolarizzazione ed E , numero di iterazioni. Come spiegato al paragrafo 3.7.4, è necessario impostare a priori il numero di tentativi dell'ottimizzatore Bayesiano (in questo caso definito pari a 500) dopodiché le combinazioni di iperparametri da valutare risulteranno campionate in funzione dell'algoritmo UCB. L'intera metodologia è stata rappresentata in Figura 7.5: la modellazione comincia con la progettazione delle miscele e l'esecuzione delle prove di laboratorio necessarie a definire le variabili di input e quelle di output; dopodiché, una rete neurale artificiale la cui struttura ed il cui funzionamento algoritmico non sono noti a priori tenterà di identificare le relazioni che funzionalmente legano input ed output attraverso il susseguirsi delle iterazioni dell'algoritmo di addestramento e l'identificazione del miglior insieme di iperparametri tramite ottimizzazione Bayesiana. Gli intervalli di ricerca di ciascun iperparametro sono stati così definiti:

- $X_L = \{1, 2, \dots, 5\}$, per il numero di strati nascosti L ;
- $X_N = \{4, \dots, 64\}$, per il numero di neuroni artificiali N in ciascuno strato nascosto;
- $X_{act} = \{ELU, ReLU, TanH\}$, per la funzione di attivazione da applicare a ciascuno strato nascosto;
- $X_\alpha = [10^{-6}, 10^{-2}]$, per il tasso di apprendimento α ;
- $X_\beta = [10^{-6}, 10^{-2}]$, per l'iperparametro di regolarizzazione β ;
- $X_E = [500, \dots, 5000]$, per il numero di iterazioni E dell'algoritmo di addestramento.

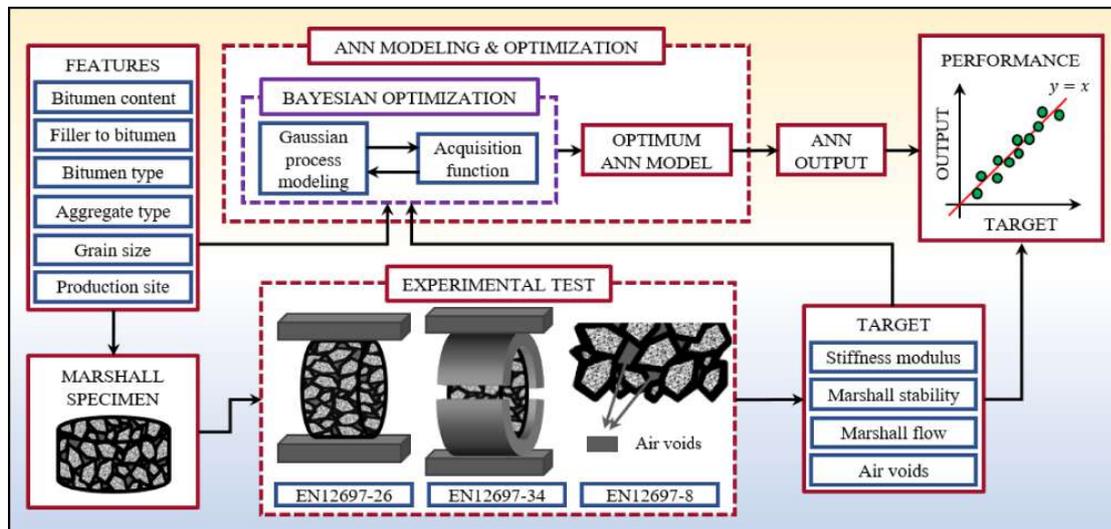


Figura 7.5: Metodologia descritta *step-by-step*.

Prima di essere utilizzate dal modello neurale, le osservazioni relative a ciascuna variabile sono state sottoposte a standardizzazione: ciascun valore è stato sottratto al rispettivo valore medio e diviso per il rispettivo valore di deviazione standard. Il codice sorgente relativo al modello qui delineato è stato scritto utilizzando il linguaggio Python ed è stato eseguito su un computer dotato di CPU Intel(R) Xeon(R) W-2125 a 4 GHz e di 32 GB di RAM. La durata di ciascun esperimento è risultata approssimativamente pari a 24 ore.

7.4 Discussione dei risultati

In Figura 7.6, è possibile osservare i valori medi del coefficiente di correlazione di Pearson determinati a seguito della procedura di convalida incrociata (R_{CV}) per ciascuna delle 500 iterazioni dell' algoritmo di ottimizzazione Bayesiana. Nello specifico, è possibile osservare come, durante le prime 350 iterazioni, i valori di R_{CV} risultino all'incirca uniformi, mostrando variazioni più marcate durante le iterazioni successive. L' algoritmo di ottimizzazione ha evidentemente individuato una regione dell'iperspazio di ricerca in cui l'errore di validazione non variava in maniera significativa ed, avendo valutato una sovra-esplorazione della zona in esame, ha deciso di focalizzare la ricerca in un'area ancora inesplorata che avrebbe potuto restituire soluzioni con prestazioni migliori (identificando tuttavia delle combinazioni peggiori). Tale risultato evidenzia come l' algoritmo UCB ricerchi il giusto compromesso tra sfruttamento ed esplorazione dello spazio degli iperparametri.

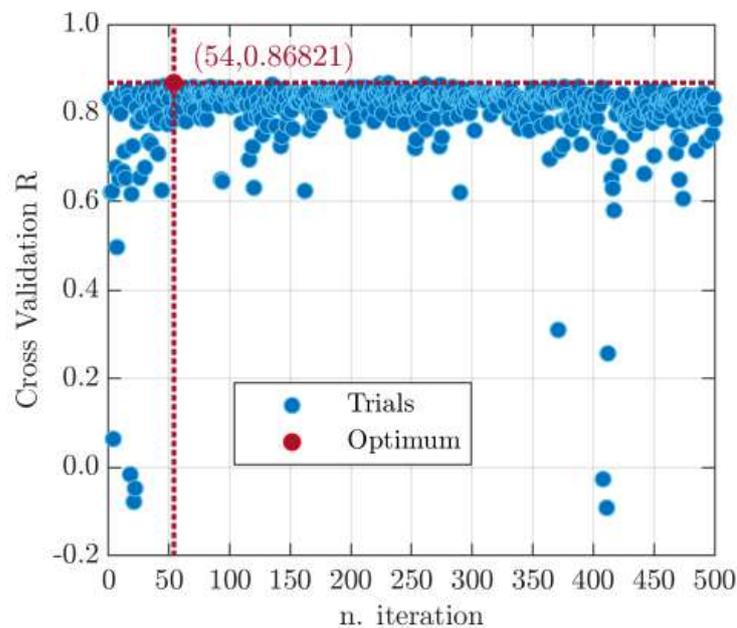


Figura 7.6: Andamento del coefficiente R_{CV} valutato durante le 500 iterazioni del BO.

Come si può notare in Figura 7.6, nonostante le notevoli dimensioni dello spazio di ricerca, il miglior modello (che restituisce un valore minimo di MSE_{CV} pari a circa 0.249 ed un valore massimo di R_{CV} pari a circa 0.868) è stato identificato all'iterazione 54. I valori ottimali di ciascun iperparametro sono risultati: L pari a 5, N pari a 37, act pari a $TanH$, E pari a 3552, α pari a 0.01 e β pari a 10^{-6} . In riferimento a tale modello, la Tabella 7.5 riporta tutti i valori di MSE e R ottenuti per ciascun gruppo (*fold*) della procedura di convalida incrociata.

Tabella 7.5: Valori di MSE e R ottenuti tramite ottimizzazione BO.

Fold	MSE_{medio}	Coeff. Corr. Pearson R				R_{medio}
		ITSM	MS	MF	AV	
0	0.219	0.837	0.866	0.842	0.964	0.877
1	0.203	0.963	0.835	0.725	0.973	0.874
2	0.254	0.872	0.836	0.799	0.917	0.856
3	0.223	0.918	0.826	0.917	0.956	0.905
4	0.346	0.841	0.731	0.834	0.912	0.829
Risultato $_{CV}$	0.249	0.886	0.819	0.824	0.944	0.868

La corrispondenza tra osservazioni predette e quelle sperimentali di target può essere osservata mediante i diagrammi di regressione riportati in Figura 7.7, in cui i dati sono suddivisi per ciascuna variabile di output e per ciascun gruppo (*fold*) della procedura di convalida incrociata. Nonostante l'elevata variabilità delle osservazioni di ciascuna variabile, che può essere giustificata considerando le diverse proprietà dei conglomerati bituminosi analizzati, il modello ottimale restituisce risultati molto incoraggianti (Figura 7.7). Come si può osservare anche in Tabella 7.5, il coefficiente di correlazione di Pearson mediato sui 5 *fold* risulta sempre superiore a 0.819, qualunque sia il parametro fisico-meccanico considerato. Le fluttuazioni osservate nell'andamento dei parametri MSE e R (seconda e settima colonna della Tabella 7.5) possono essere attribuite alla diversa distribuzione delle osservazioni di addestramento e di test che caratterizzano ogni gruppo. Tale risultato mostra come una suddivisione fissa della banca dati (nota come *hold-out procedure*) possa provocare una valutazione errata dell'errore commesso in fase predittiva che può essere peggiore (ad esempio $R_4 = 0.829$) oppure migliore (ad esempio $R_3 = 0.906$) rispetto alla più probabile condizione rappresentata da $R_{CV} = 0.868$. Infine, la Figura 7.8 mostra un confronto tra i target sperimentali e gli output previsti dal modello per ciascuno dei quattro parametri fisico-meccanici analizzati. A titolo esemplificativo, i risultati riportati riguardano il *fold* numero 4, ovvero quello che ha restituito i valori "peggiori" in termini di MSE_{medio} ed R_{medio} . Anche in questo caso, infatti, i valori predetti per ciascuna variabile di output sono molto prossimi ai valori osservati sperimentalmente. Tale risultato risulta di notevole rilevanza da un punto di vista ingegneristico poiché dimostra come le reti neurali artificiali possano costituire modelli particolarmente accurati per la previsione della risposta meccanica e delle proprietà fisiche (anche in maniera simultanea) di conglomerati bituminosi anche molto differenti tra loro in termini di composizione. Sebbene la procedura di modellazione presentata in questo capitolo sia concettualmente e computazionalmente più complessa di quelle convenzionali presenti in letteratura, essa supera alcuni dei problemi insiti nei metodi tradizionali (come le valutazioni distorte delle prestazioni del modello e la selezione

sub-ottimale degli iperparametri) o nei *toolbox* più frequentemente utilizzati (come ad esempio il MATLAB®ANN Toolbox), come emerge dall'analisi dei risultati riportati.

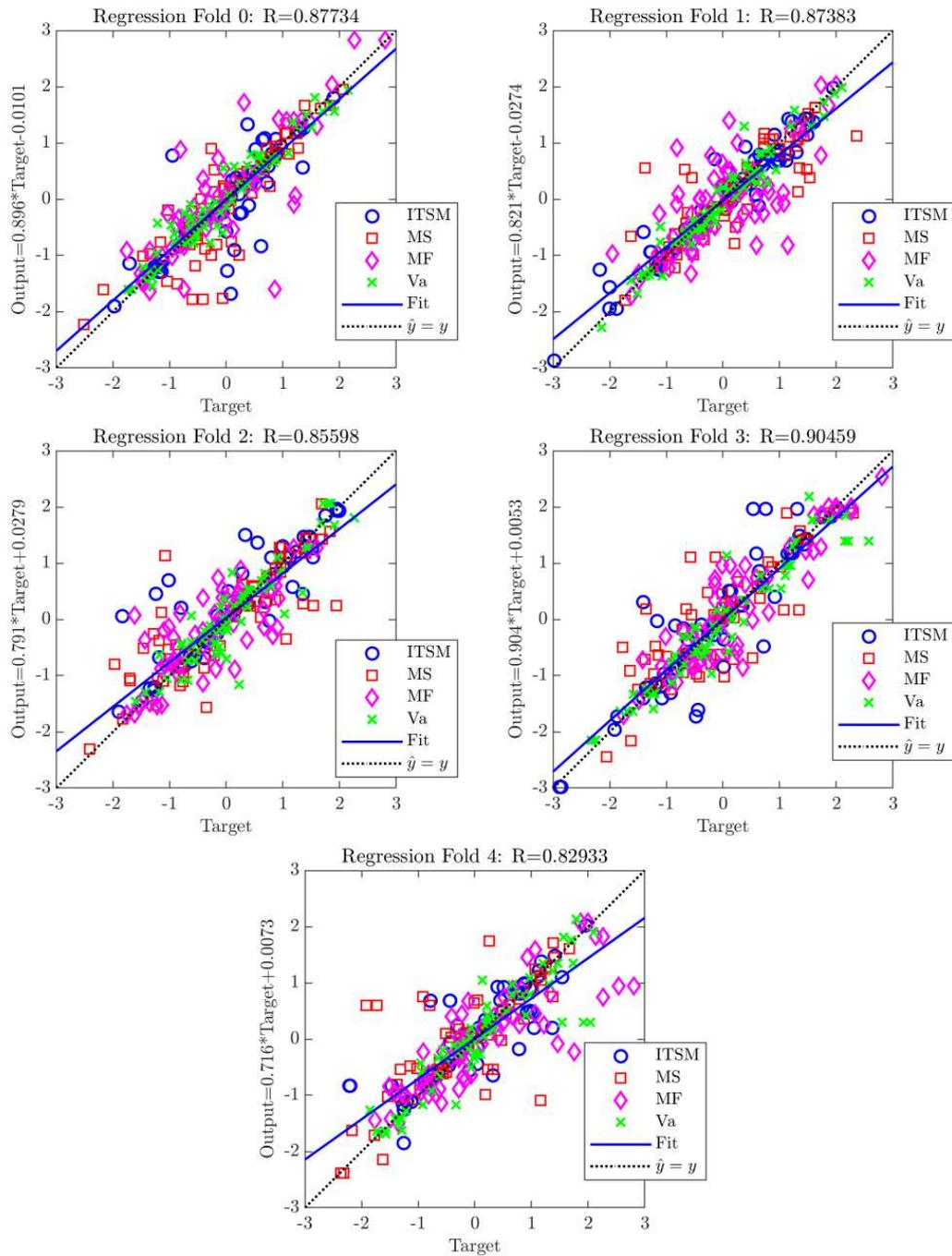


Figura 7.7: Risultati della regressione riferiti ai 5 gruppi della convalida incrociata.

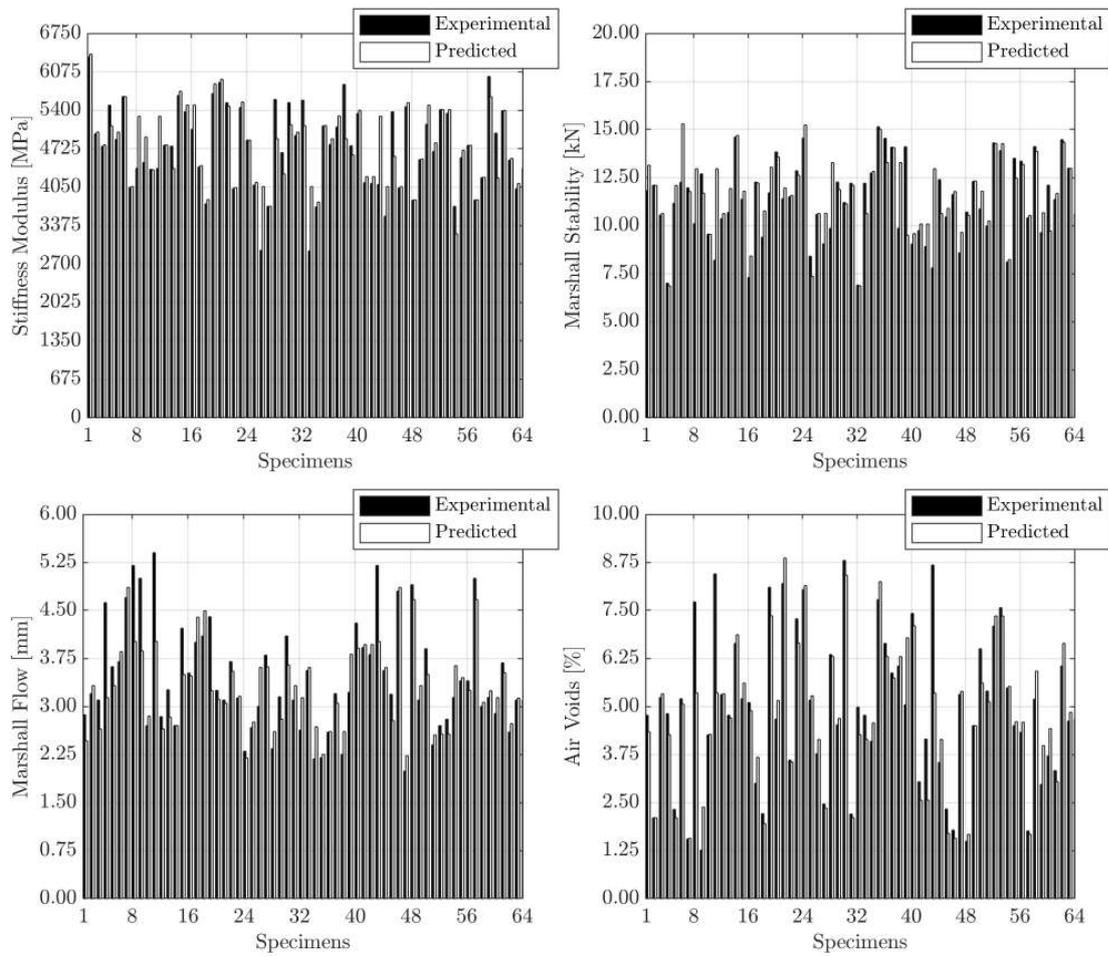


Figura 7.8: Confronto tra target sperimentali ed output previsti relativo al *fold* 4.

7.5 Conclusioni

L'obiettivo principale di questo capitolo è stato quello di discutere ed implementare una procedura semi-automatica ed imparziale per la selezione del modello neurale più affidabile, in grado di superare i limiti degli approcci convenzionali, assegnando un determinato problema di modellazione predittiva. In particolare, l'attenzione è stata rivolta alla previsione simultanea delle proprietà volumetriche e delle caratteristiche meccaniche di miscele bituminose preparate utilizzando diverse tipologie di bitume ed aggregati, molteplici contenuti di legante e diverse granulometrie. Ciò al fine di supportare la fase di progettazione delle miscele, fornendo stime numeriche dei parametri investigati e riducendo la necessità di effettuare ulteriori onerose prove di laboratorio. A tale scopo, un'estesa campagna sperimentale è stata eseguita presso il Laboratorio di Ingegneria Autostradale dell'Università Aristotele di Salonicco. Per caratterizzare le miscele sono stati utilizzati test del modulo di rigidità [UNI EN 12697-26, 2018], dei parametri Marshall [UNI EN 12697-34, 2020] e di caratterizzazione volumetrica [UNI EN 12697-8, 2019]. Poiché questi sono alcuni dei parametri maggiormente utilizzati nella progettazione dei conglomerati bituminosi per pavimentazioni stradali, tali proprietà meccaniche e volumetriche sono state correlate al contenuto di bitume, al rapporto tra filler e bitume e ad una variabile categorica che rifletta indirettamente la tipologia di miscela. Gli iperparametri ottimali del modello di apprendimento automatico che minimizzino gli errori commessi in fase predittiva sono stati determinati utilizzando l'ottimizzazione Bayesiana e la convalida incrociata. I principali risultati raggiunti possono essere riassunti come segue:

- Al fine di eseguire una corretta modellazione neurale, è necessario valutare le molteplici strutture di rete risultanti dalla selezione di diversi valori degli iperparametri del modello. La procedura sviluppata in questo capitolo ha permesso di superare le limitazioni dei convenzionali *toolbox* ANN;
- L'approccio dotato di convalida incrociata restituisce risultati maggiormente attendibili rispetto ad un semplice partizionamento casuale della banca dati: infatti, se la procedura fosse stata basata su un singolo partizionamento casuale, sarebbero stato possibile ottenere risultati differenti, peggiori (come nel caso del *fold* 4) o migliori (come nel caso del *fold* 3) rispetto alla condizione più probabile rappresentata dal risultato della *cross-validation*;
- L'algoritmo di ottimizzazione Bayesiana ha dimostrato di riuscire a risolvere con successo lo spinoso problema relativo alla corretta impostazione degli iperparametri del modello: esso ha infatti consentito di individuare la soluzione ottimale in termini di configurazione algoritmica e strutturale della rete neurale artificiale dopo appena 54 iterazioni. Il tratto distintivo di tale tecnica risiede nella sua capacità di tenere conto delle valutazioni precedenti al fine di sfruttare ed esplorare lo spazio degli iperparametri. Tuttavia, è necessario ricordare che i risultati della procedura BO risultano vincolati ai limiti imposti in termini di variabilità degli iperparametri;
- La struttura di rete neurale artificiale maggiormente idonea a modellare le osservazioni sperimentali in esame è definita da 5 strati nascosti, 37 neuroni per ciascuno strato, un algoritmo di addestramento Ranger, un funzione di trasferimento tangente iperbolica, un

tasso di apprendimento pari a 0.01 ed un parametro di regolarizzazione dei pesi pari a 10^{-6} e ha restituito valori di R_{CV} e MSE_{CV} rispettivamente pari a 0.868 e 0.249.

La metodologia messa a punto è stata descritta nel dettaglio al fine di rendere la procedura seguita riproducibile dal lettore. Tuttavia, è opportuno ricordare che, sebbene la banca dati sia particolarmente estesa, il modello proposto è ottimizzato ed applicabile esclusivamente in relazione alle specifiche tipologie di aggregato, bitume e struttura della miscela considerati. Un interessante sviluppo futuro potrebbe essere rappresentato dallo studio dell'effetto che diverse granulometrie degli aggregati esercitano a parità di uno specifico contenuto di bitume. Un ulteriore approfondimento potrebbe riguardare l'integrazione di innovative variabili relative alle prestazioni della pavimentazione stradale, come la resistenza alla fatica ed alle deformazioni permanenti, in sostituzione delle caratteristiche meccaniche delle miscele. Tale transizione rappresenterebbe un ulteriore progresso a favore di una progettazione dei conglomerati basata sulle prestazioni.

Capitolo 8

Influenza di Filler Sostenibili sul Comportamento Meccanico di Miscele Bituminose

” *Sustainable development is the pathway to the future we want for all. It offers a framework to generate economic growth, achieve social justice, exercise environmental stewardship and strengthen governance.*

— **Ban Ki-moon**

8.1 Introduzione

Una pavimentazione stradale flessibile in esercizio è solitamente soggetta a processi di deterioramento indotti da traffico, acqua, invecchiamento ecc. [Wang et al., 2017]. Per questo motivo, è fondamentale che il conglomerato bituminoso progettato garantisca prestazioni fisico-meccaniche adeguate. Come descritto nel Capitolo 2, le miscele bituminose sono per lo più costituite da aggregato grossolano e bitume; tuttavia, è necessario aggiungere e bilanciare anche un'adeguata percentuale di filler. Infatti, una corretta quantità di filler contribuisce sia a ridurre la porosità della miscela sia a migliorare le prestazioni complessive in termini di resistenza ai fenomeni di ormaiamento e rigonfiamento [Yilmaz et al., 2011, Chen et al., 2011, Veytskin et al., 2015, Dash and Panda, 2018, Miró et al., 2017, Raposeiras et al., 2018]. Tuttavia, una quantità eccessiva di filler può comportare la diminuzione del contenuto di aggregato grossolano, inducendo una conseguente riduzione della resistenza meccanica complessiva della miscela [Xue et al., 2009]. Nel corso degli anni, i ricercatori hanno investigato gli effetti che i diversi riempitivi minerali generavano nei conglomerati in cui venivano utilizzati [Ogundipe, 2016, Lesueur et al., 2013, Rashwan, 2016, Chrismer and Durham, 2010, Abdel-Wahed and Rashwan, 2016], giungendo a stabilire che il cemento produceva i più significativi incrementi nelle prestazioni meccaniche delle miscele bituminose [Mazzoni et al., 2019]. Tuttavia, la crescente preoccupazione per

i temi legati alla sostenibilità ambientale ha evidenziato le elevate emissioni di CO₂ inevitabilmente associate alla produzione del cemento. Per questo motivo, dunque, è necessario identificare una valida alternativa che sia al tempo stesso eco-sostenibile e meccanicamente valida [Schneider et al., 2011]. A tal fine, risulta necessaria una caratterizzazione completa del filler individuato come potenziale efficace sostituto, sia esso riciclato oppure derivante da rifiuti industriali. Occorre infatti determinare i parametri più critici che potrebbero influenzare le globali prestazioni meccaniche evidenziate dal conglomerato bituminoso, ovvero: forma e dimensione delle particelle, peso specifico, area superficiale, composizione mineralogica e presenza di materiale fine potenzialmente dannoso [Wang et al., 2011, Tiwari and Satyam, 2021].

In questo specifico contesto si colloca il lavoro di molteplici ricercatori di tutto il mondo, i quali hanno analizzato le variazioni nel comportamento meccanico dei conglomerati bituminosi indotte dall'utilizzo di alcuni potenziali filler alternativi come ad esempio: la polvere di scorie di rame [Modarres and Alinia Bengar, 2019], il residuo di bauxite (noto anche come fango rosso) [Choudhary et al., 2022], la cenere di frutti di palma vuoti [Lukjan et al., 2022] e la cenere di buccia di caffè [Tessema et al., 2022]. Recentemente, la risposta fisico-meccanica delle miscele di conglomerato bituminoso preparate con polvere di marmo si è dimostrata tecnicamente valida e preferibile a quella raggiunta utilizzando filler ottenuti da scorie di acciaio o granito [Awed et al., 2022]. In linea generale, si è constatato che basse percentuali di SiO₂, abbinate ad elevate percentuali di CaO, determinano una maggiore resistenza all'acqua ed una riduzione dei danni indotti dall'umidità [Hamedi et al., 2021]. Pertanto, la possibilità di ottenere prestazioni meccaniche equivalenti, o addirittura leggermente migliori, sostituendo i riempitivi tradizionali con materiali di scarto e la contemporanea riduzione dei problemi di smaltimento dei rifiuti rappresentano una delle moderne frontiere relative al tema della progettazione eco-sostenibile delle miscele bituminose.

Lo studio descritto in questo capitolo si è concentrato sull'analisi di conglomerati bituminosi preparati con molteplici materiali utilizzati come filler alternativi in sostituzione del tradizionale cemento Portland (*Ordinary Portland Cement* - OPC), utilizzato come prodotto di riferimento per i successivi confronti. Le potenziali soluzioni tecnologiche alternative prevedevano miscele messe a punto utilizzando riempitivi derivanti da cenere di lolla di riso (*Rice Husk Ash* - RHA), polvere di mattoni (*Brick Dust* - BD), ceneri volanti (*Fly Ash* - FA), polvere di calcare (*Limestone Dust* - LD) e fumo di silice (*Silica Fume* - SF). Una completa caratterizzazione meccanica oltre che un'approfondita valutazione comparativa è stata effettuata per ciascuna alternativa menzionata. Inoltre, sui conglomerati bituminosi confezionati, sono state eseguite prove di stabilità e scorrimento Marshall, di resistenza alla trazione indiretta, di sensibilità all'acqua e di resistenza all'abrasione. Sono stati presi in considerazione quattro diversi contenuti percentuali di filler ritenuti significativi per la tecnologia stradale, tutti compresi tra il 4% e l'8.5% in volume della miscela, con incrementi fissati dell'1.5%.

Le osservazioni derivanti dalla campagna sperimentale descritta sono state elaborate tramite un modello di apprendimento automatico basato sul moderno approccio denominato Categorical Boosting (o CatBoost). In linea generale, negli ultimi decenni, algoritmi basati su macchine a vettori di supporto, reti neurali artificiali ed alberi decisionali stanno ottenendo ampi consensi nella comunità scientifica poiché risultano in grado di prevedere in maniera attendibile il comportamento meccanico dei conglomerati bituminosi, in maniera più rapida ed accurata del-

le convenzionali metodologie empiriche e/o statistiche [Sholevar et al., 2022, Hou et al., 2021]. Sebbene i modelli basati sul *Machine Learning* trascurino la natura fisica del problema investigato, questi sono in grado di fornire un'alternativa affidabile alle avanzate equazioni costitutive relative alla meccanica dei materiali [Pasetto and Baldo, 2016, de Oliveira Junior and de Farias, 2020, Cao et al., 2020, Sun et al., 2022, Pasetto and Baldo, 2015]. Per questo motivo, nel corso degli anni, si sono affermati diversi approcci ML di successo [Baldo et al., 2022b, Baldo et al., 2022a, Upadhyay et al., 2023] che hanno permesso di prevedere il comportamento meccanico dei conglomerati bituminosi in termini di resistenza a fatica [Behbahani et al., 2020], a trazione indiretta [Gilani et al., 2020] e di indice di rugosità internazionale (*International Roughness Index* - IRI) [Marcelino et al., 2021, Zeiada et al., 2020]. Per questo motivo, si è deciso di valutare la possibilità di sviluppare un approccio ML basato sull'algoritmo CatBoost, in grado di prevedere la risposta meccanica dei conglomerati bituminosi preparati utilizzando filler alternativi sulla base di poche e semplici variabili compositive. Per compensare il numero relativamente ridotto di osservazioni derivanti dalle indagini di laboratorio, una tecnica innovativa di generazione sintetica dei dati (Data Augmentation) è stata implementata al fine di espandere la dimensione della banca dati di partenza, consentendo un adeguato addestramento del modello.

La peculiarità del presente studio risiede nel riciclo di materiali di scarto come filler alternativi nei conglomerati bituminosi, aspetto che consente la contemporanea riduzione dei problemi di smaltimento rifiuti in discarica. Inoltre, il confronto diretto delle prestazioni meccaniche evidenziate dalle miscele preparate con otto diversi riempitivi alternativi e l'utilizzo di moderne tecniche di apprendimento automatico, rappresentano gli aspetti maggiormente innovativi di questo specifico lavoro di ricerca. Il risultato complessivo è rappresentato da un approccio integrato che, al fine di stimare la risposta fisico-meccanica delle miscele preparate con filler selezionati, ha previsto dapprima un'approfondita indagine sperimentale seguita dallo sviluppo e dall'implementazione di un adeguato modello previsionale.

8.2 Campagna Sperimentale

8.2.1 Materiali

Gli aggregati a spigoli vivi utilizzati per la preparazione delle miscele oggetto di questo capitolo sono stati derivati dalla frantumazione del quarzo e sono stati forniti dalla Safew Tech Systems, situata a Indore, India. Per una loro completa caratterizzazione, molteplici prove di laboratorio sono state eseguite. I parametri obiettivo, le procedure seguite, i risultati ottenuti ed i limiti imposti dalla normativa indiana (conosciuta anche come MoRTH - *Indian Ministry of Road Transport and Highways*) sono stati riportati nel dettaglio in Tabella 8.1.

Successivamente, la curva granulometrica progettata è stata confrontata con il fuso granulometrico specificato dalla normativa MoRTH relativamente alla pavimentazioni stradali flessibili: i risultati sono stati graficamente rappresentati in Figura 8.1.

Relativamente al bitume, questo è stato fornito da una delle principali aziende private indiane produttrici di bitume. Si è trattato di un convenzionale VG-30 di cui viene fornita una descrizione dettagliata in Tabella 8.2.

Tabella 8.1: Proprietà dell'aggregato.

Parametro di prova	Risultato	Limiti MoRTH	Normativa di riferimento
Pulizia (%)	3	(max) 5	IS 2386 Parte I
Peso specifico (g/cm ³)	2.68	2÷3	IS 2386 Parte III
Coefficiente Los Angeles (%)	10.6	(max) 35	IS 2386 Parte IV
Resistenza all'effetto del solfato di sodio (%)	3.4	(max) 12	IS 2386 Parte V
Resistenza all'effetto del solfato di magnesio (%)	3.7	(max) 18	IS 2386 Parte V
Indice di appiattimento (%)		(max) 35	IS 2386 Parte I
20 mm	27.93		
10 mm	32.13		
Resistenza agli urti (%)		(max) 27	IS 2386 Parte IV
20 mm	4.15		
10 mm	5.91		
Assorbimento d'acqua (%)	1.67	(max) 2	IS 2386 Parte III

Infine, in termini di filler, è stato utilizzato il tradizionale cemento Portland (OPC) come riferimento dal momento che rappresenta il materiale più diffusamente impiegato per la realizzazione di conglomerati bituminosi per strati di base di pavimentazioni stradali soggette a carichi pesanti in India. Sette diverse alternative all'OPC, derivanti dal riciclo di materiali di scarto, sono state analizzate ed un riepilogo dell'analisi comparativa è stato riportato in Tabella 8.3. Le proprietà fisico-chimiche riportate hanno preso in considerazione il peso specifico, il valore di blu di metilene (*Methylene Blue Value* - MBV), il modulo di finezza (*Fineness Modulus* - FM), l'area superficiale, il pH, il contenuto percentuale dei principali ossidi e la forma dei grani.

8.2.2 Indagini di laboratorio

Per valutare esclusivamente l'impatto fisico-meccanico dei filler selezionati sulle miscele di riferimento, alcune variabili volumetriche sono state mantenute costanti. Infatti, il tipo di aggregato, la curva granulometrica ed il tipo di bitume sono rimasti invariati. Inoltre, è stato determinato il contenuto ottimale di bitume per ciascun conglomerato bituminoso. Come descritto in precedenza, sono stati selezionati otto diversi filler in quattro quantitativi diversi (4,0, 5,5, 7,0 e 8,5% in volume percentuale della miscela), per un totale di 32 miscele analizzate. La forma delle particelle di ciascun filler è stata riportata nella Tabella 8.3 mentre una rappresentazione grafica di RHA, BD, MD, SD, FA, OPC, LD e SF è stata fornita in Figura 8.2. I diversi riempitivi e gli aggregati sono stati preriscaldati singolarmente a 150 °C. Successivamente, la quantità necessaria di aggregati e filler è stata miscelata a 170 °C in un mescolatore planetario alla velocità di 140 ± 5 giri/min. Le rispettive quantità di bitume sono state poi aggiunte secondo gli standard MoRTH per preparare le miscele. Le proprietà volumetriche dei conglomerati bituminosi hanno un impatto significativo sulla resistenza meccanica e sulla durabilità a lungo termine delle pavimentazioni stradali. Per questo motivo, sono state misurate le percentuali di vuoti d'aria

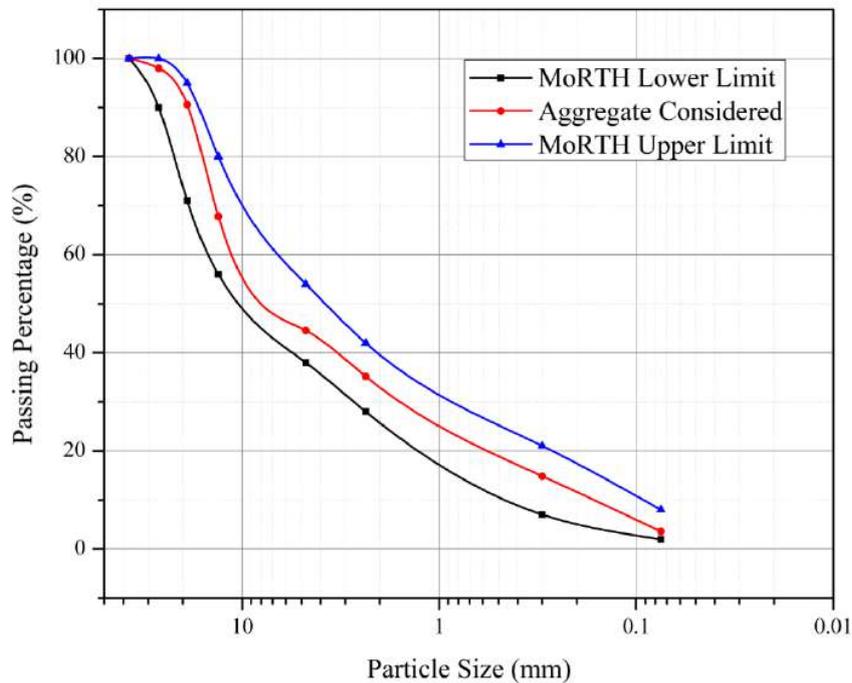


Figura 8.1: Curva granulometrica dell'aggregato progettata secondo requisiti MoRTH.

(*Air Voids - AV*), di vuoti nell'aggregato minerale (*Voids in the Mineral Aggregate - VMA*) e di vuoti riempiti di bitume (*Voids Filled with Bitumen - VFB*) in conformità agli standard MoRTH. Per preparare i provini per le successive indagini sperimentali, è stato seguito il metodo Marshall. Sebbene questa tecnica abbia dimostrato di condurre a proprietà volumetriche leggermente diverse da quelle successivamente osservate nelle pavimentazioni stradali [Sousa et al., 1995], è ancora ampiamente utilizzata nei laboratori stradali [Xue et al., 2009, Pasandín and Pérez, 2015, Al-ammari et al., 2019, Ozgan, 2011, Zavrtnik et al., 2016] ed è talvolta preferita al metodo Superpave [Cominsky et al., 1994] per la sua semplicità ed economicità. Prima di effettuare i test Marshall, ogni provino (67 ± 3 mm di altezza e 102 mm di diametro) è stato compattato con 75 colpi di martello Marshall su ciascun lato. Successivamente, è stato raffreddato in aria per 4 ore e infine tenuto per 35 ± 5 minuti in un bagno d'acqua alla temperatura di 60 ± 1 °C. Una cella di carico da 100 kN è stata utilizzata insieme ad un trasduttore di spostamento lineare variabile da 50 mm per ottenere il valore massimo della resistenza al carico e la corrispondente deformazione per una velocità costante di deformazione pari a 50,4 mm/min. I risultati ottenuti da questa indagine di laboratorio sono di seguito indicati come stabilità Marshall (*Marshall Stability - MS*), scorrimento (*Marshall Flow - MF*) e quoziente (*Marshall Quotient - MQ*) secondo quanto previsto dalla normativa [ASTM D6927-15, 2015].

MQ è definito come il rapporto tra MS e MF ed è empiricamente correlato alle prestazioni delle miscele bituminose in termini di resistenza all'ormaiamento [Zoorob and Suparma, 2000]. Quanto più alto è il valore di MQ, tanto migliore sarà la resistenza della miscela alle deformazioni



Figura 8.2: Esempi di RHA (a), BD (b), MD (c), SD (d), FA (e), OPC (f), LD (g) e SF (h).

Tabella 8.2: Proprietà del bitume.

Parametro di prova	Risultato	Limiti MoRTH	Normativa di riferimento
Viscosità assoluta a 60°C (Poises)	2855	2400 ÷ 3600	IS 1206 Parte 2
Viscosità cinematica a 135°C (cSt)	392	(min) 350	IS 1206 Parte 3
Punto di infiammabilità COC (°C)	304	(min) 250	IS 1448 Parte 69
Penetrazione a 25°C (dmm)	49	(min) 45	IS 1203
Punto di rammollimento (°C)	48	(min) 47	IS 1205
Solubilità in solventi organici (% in massa)	99.45	(min) 99	IS 1216
Rapporto di viscosità a 60°C	1.3	(max) 4.0	IS 1206 Parte 2
Duttilità a 25°C dopo TFOT (cm)	75	(min) 40	IS 1208
Peso specifico (g/cm ³)	0.987	0.97 ÷ 1.02	IS 1202

da *creep* indotte dal traffico. Inoltre, per tutte le miscele sono state eseguite prove di resistenza a trazione indiretta ITS secondo la normativa [ASTM D6931-17, 2017]. Per valutare la sensibilità all'acqua è stato eseguito un test di Lottman modificato. I valori ITS sono stati determinati prima e dopo il condizionamento (in conformità con la normativa [AASHTO T283, 2022]), e il rapporto tra di essi ha determinato il valore del rapporto di resistenza a trazione indiretta (*Indirect Tensile Strength Ratio* - ITS_R) indicativo della sensibilità all'acqua. Infine, per tenere conto del potenziale degrado della miscela, è stata eseguita una prova Cantabro di resistenza all'abrasione secondo la specifica [Materials and Test Division Texas Department of Transportation, 2019]. I provini di 101 ± 1,5 mm di diametro e 50,8 ± 1,5 mm di spessore sono stati compattati con il martello Marshall con 50 colpi per ogni lato ed è stato determinato il peso iniziale di ciascun provino a 25 °C (m_i). Dopo 24 ore, il provino è stato posizionato in una macchina di prova Los Angeles, senza le cariche abrasive. La prova è stata eseguita alla velocità di 32 giri/minuto, per un totale di 300 giri. Il materiale frammentato è stato scartato ed è quindi stato determinato il peso finale del provino (m_f). La perdita Cantabro (*Cantabro Loss* - CL) è stata calcolata secondo l'equazione:

$$CL = \frac{m_i - m_f}{m_f} \times 100 \quad (8.2.1)$$

Tabella 8.3: Proprietà dei filler analizzati.

Parametro di prova	Tipologia di filler							
	RHA	BD	MD	SD	FA	OPC	LD	SF
Peso specifico (g/cm ³)	2.02	2.56	2.69	2.69	2.32	3.04	2.65	2.20
MBV (g/kg)	4.72	6.25	4.45	3.67	3.86	3.00	3.75	3.85
FM	3.21	5.17	2.12	5.38	3.77	4.96	3.03	1.96
Area superficiale (m ² /g)	2.31	2.69	4.37	2.70	2.19	1.75	2.70	16.45
pH	10.86	8.67	8.50	12.57	7.30	12.90	10.22	6.98
SiO ₂ (%)	89.67	39.55	0.60	82.37	48.24	21.43	0.48	93.50
CaO (%)	1.88	12.88	55.60	2.79	13.40	66.58	96.57	0.89
Al ₂ O ₃ (%)	1.62	15.71	0.40	8.23	24.15	3.01	0.41	0.08
MgO (%)	0.97	3.29	0.10	1.47	1.46	1.39	0.46	0.82
Fe ₂ O ₃ (%)	1.06	14.05	0.20	5.27	6.48	4.68	0.32	0.50
Forma dei grani	Alveolata	Particelle sub-angolari	Particelle sub-angolari	Particelle angolari	Arrotondata	Particelle granulari/ sub-angolari	Particelle granulari	Sferica

8.3 Modellazione predittiva con approccio CatBoost

Le informazioni provenienti dalle indagini di laboratorio sono state elaborate mediante una moderna tecnica di ML denominata CatBoost. Come ampiamente descritto nel paragrafo 4.4.8, si tratta di un algoritmo innovativo che deriva dagli alberi decisionali basati sulla tecnica *Gradient Boosting*, ma che risolve i principali inconvenienti comunemente indicati come *prediction shift* (o scostamento previsionale) e *target leakage* [Prokhorenkova et al., 2018]. Combinando l'implementazione delle tabelle decisionali [Lou and Obukhov, 2017] e dell'*Ordered Boosting* [Prokhorenkova et al., 2018], CatBoost definisce un'architettura ad albero più efficiente e bilanciata che migliora le prestazioni degli attuali algoritmi basati su alberi decisionali [Bentéjac et al., 2021] come XGBoost [Chen and Guestrin, 2016] e LightGBM [Ke et al., 2017]. Inoltre, CatBoost consente di elaborare e analizzare automaticamente gli insiemi di dati contenenti variabili di tipo categorico.

Per ottimizzare le prestazioni del modello CatBoost, i suoi iperparametri necessitano di un'attenta calibrazione. Per identificare i valori migliori in termini di numero di alberi da costruire (*number of iterations*), di profondità massima di tali alberi (*max depth*) e del tasso di apprendimento (*learning rate*), è stata implementata un'estesa ricerca a griglia (sintetizzata in Tabella 8.4). Per evitare fenomeni di *overfitting*, oltre alla condizione relativa alla massima profondità dell'albero, sono stati implementati una procedura di convalida incrociata k-fold [Baldo et al., 2021] ed un rilevatore di *overfitting* [James et al., 2013]. In questo caso particolare, il valore di k è stato impostato pari a 4, in modo che ogni *fold* contenesse lo stesso numero di osservazioni. Il numero di iterazioni del rilevatore di *overfitting*, invece, è stato impostato al valore predefinito di 20, in base a quanto suggerito dalla letteratura scientifica di riferimento [Rahman et al., 2019]. Il modello migliore è stato identificato minimizzando il valore della funzione di perdita. In questo caso, la funzione maggiormente idonea è stata individuata nel MultiRMSE, poiché l'obiettivo della modellazione consisteva nella previsione simultanea di cinque diversi parametri (MS, MQ, ITS, ITSR e CL) sulla base di tre semplici input, ovvero: una variabile categorica identificativa del tipo di filler (CV), il contenuto percentuale di filler (Fc) e la percentuale dei vuoti d'aria (AV).

Tabella 8.4: Riepilogo della ricerca a griglia.

Variabile	Griglia	Valore selezionato
Numero di alberi	100, 250, 500, 1000, 2000	1000
Profondità massima	1, 2, 3, 4, 5	3
Tasso di apprendimento	0.5, 0.1, 0.05, 0.01, 0.005	0.1
k-fold CV		4
Rilevatore overfitting		20
Funzione di perdita		MultiRMSE

Identificando con D il numero complessivo di variabili di output, con N il numero di osservazioni contenute all'interno del vettore di test, con y_{T_i} l' i -esimo target e con y_{P_i} l' i -esima predizione, l'espressione analitica del MultiRMSE risulta:

$$\text{MultiRMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \sum_{d=1}^D (y_{T_{i,d}} - y_{P_{i,d}})^2} \quad (8.3.1)$$

Per misurare con precisione l'attendibilità delle previsioni, sono state implementate sei diverse metriche di prestazione: MAE, MAPE, MSE, RMSE, R e R², la cui espressione analitica è riportata nel paragrafo 3.8.2, insieme al significato statistico. L'intera metodologia è stata implementata in Python 3.8.5.

8.3.1 Data Augmentation

Le prestazioni di un modello basato sull'apprendimento automatico dipendono fortemente dalla qualità, dalla quantità e dal livello di significatività delle osservazioni di addestramento. Le dimensioni ridotte delle banche dati rappresentano una delle problematiche cruciali nello sviluppo dei modelli predittivi, poiché la raccolta di una quantità considerevole di dati richiede spesso tempi molto lunghi e costi molto elevati. Le tecniche di *Data Augmentation* aumentano artificialmente la quantità di dati disponibili generando nuovi dati sintetici sulla base delle osservazioni sperimentali raccolte. Tuttavia, occorre soddisfare una condizione fondamentale: i dati sintetici (o aumentati) devono mantenere il significato intrinseco originale dei dati sperimentali, senza alterarlo in alcun modo [Oh et al., 2020]. In linea generale, nel settore dell'ingegneria delle pavimentazioni, si riscontrano spesso insiemi di dati di piccole dimensioni, poiché la maggior parte delle indagini di laboratorio è costosa e richiede molto tempo. Per questo motivo, le analisi del comportamento meccanico sono spesso limitate ad un numero ridotto di campioni e a poche varianti di una determinata miscela di conglomerato bituminoso. In questo studio, per stimare valori sconosciuti sulla base dei dati noti, è stata implementata la tecnica di interpolazione cubica Makima [Akima, 1974].

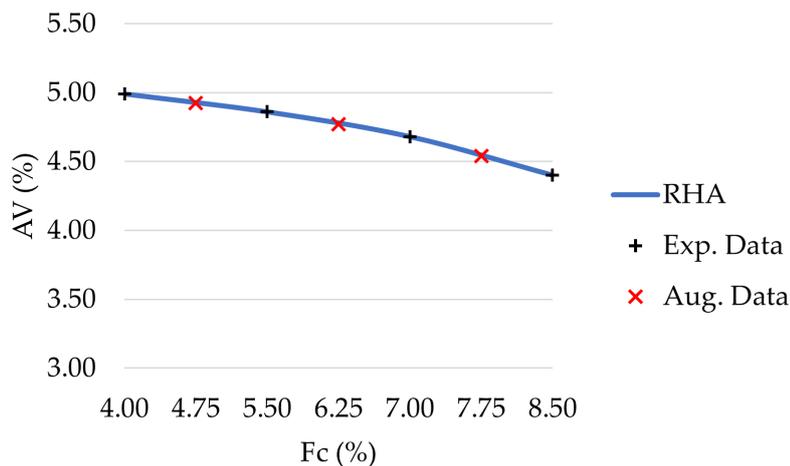


Figura 8.3: Dati sperimentali (in nero) e dati sintetici (in rosso) per le miscele RHA.

Si tratta di un miglioramento dell'algoritmo originale di Akima [Akima, 1970], e consiste nella produzione di polinomi a tratti con derivate del primo ordine continue, eseguendo un'interpolazione cubica. In questo modo, è possibile considerare anche problemi altamente non lineari e con dati oscillanti, senza però incorrere nelle eccessive oscillazioni locali provocate dall'interpolazione polinomiale di terzo grado o dalla spline cubica. Conformemente alle indicazioni di letteratura [Oh et al., 2020], la quantità di dati originali non è stata superata da quella dei dati sintetici. Pertanto, è stato implementato un algoritmo Makima volto a determinare il valore delle variabili tra due successivi dosaggi di filler investigati, ottenendo 3 osservazioni sintetiche per ciascuna curva. Tale operazione è stata eseguita per ciascuno degli otto filler considerati, fornendo 24 osservazioni aggiuntive per ciascuna variabile. Dunque, la dimensione del set di dati è stata quasi raddoppiata, passando da 32 a 56 osservazioni per ogni variabile registrata. Un diagramma esemplificativo relativo alla variabile vuoti d'aria (AV) delle miscele preparate con filler RHA è stato realizzato ed è riportato in Figura 8.3.

8.4 Discussione dei risultati

8.4.1 Risultati di laboratorio

I risultati dell'indagine di laboratorio sono stati illustrati nelle Figure da 8.4 a 8.11, in cui i limiti inferiori e quelli superiori previsti dalla normativa MoRTH per ciascun parametro sono indicati (quando specificati) rispettivamente con curve rosse e verdi. Nello specifico, l'andamento della percentuale di vuoti d'aria al variare del contenuto di filler può essere osservato nella Figura 8.4 per conglomerati bituminosi preparati con filler convenzionali e di recupero. Come previsto, un aumento del contenuto di filler ha portato a una corrispondente diminuzione dei vuoti d'aria. Tutti i tipi di miscela hanno soddisfatto i requisiti degli standard MoRTH, con valori di AV compresi nell'intervallo 3%÷5%. Tuttavia, a parità di contenuto di filler, le miscele LD e SF hanno mostrato un contenuto di vuoti d'aria inferiore rispetto alle miscele OPC, con differenze che vanno dallo 0,05% allo 0,26%.

VMA è definito come il volume dello spazio vuoto intergranulare tra gli aggregati della miscela compattata, compresi i vuoti d'aria e il volume di bitume non assorbito. Come si può osservare nella Figura 8.5, le miscele LD e SF hanno restituito valori di VMA inferiori rispetto alle miscele OPC, con differenze che vanno dallo 0,13% allo 0,41%. Considerando un contenuto di filler pari all'8,5%, le miscele preparate con BD, MD, SD, RHA, FA, OPC, LD e SF hanno rispettivamente restituito valori di VMA pari a 16,41%, 16,31%, 16,28%, 16,24%, 15,99%, 15,87%, 15,74% e 15,52%.

La percentuale di VMA riempita di bitume definisce VFB. Si può notare come un aumento del contenuto di filler abbia condotto ad un corrispondente aumento della percentuale di vuoti riempiti di bitume. In accordo con i valori di AV e VMA precedentemente descritti, la sostituzione dell'OPC con LD o SF ha leggermente aumentato il valore di VFB nelle miscele corrispondenti (Figura 8.6). Le differenze si sono attestate tra lo 0,15% e l'1,02%. Tuttavia, tutti i tipi di miscela hanno soddisfatto i requisiti previsti dagli standard MoRTH, con valori di VFB compresi nell'intervallo 65%÷85% [Ministry of Road Transport and Highways, 2013]. Considerando un contenuto di filler pari all'8,5%, le miscele preparate con FA, SD, MD, BD

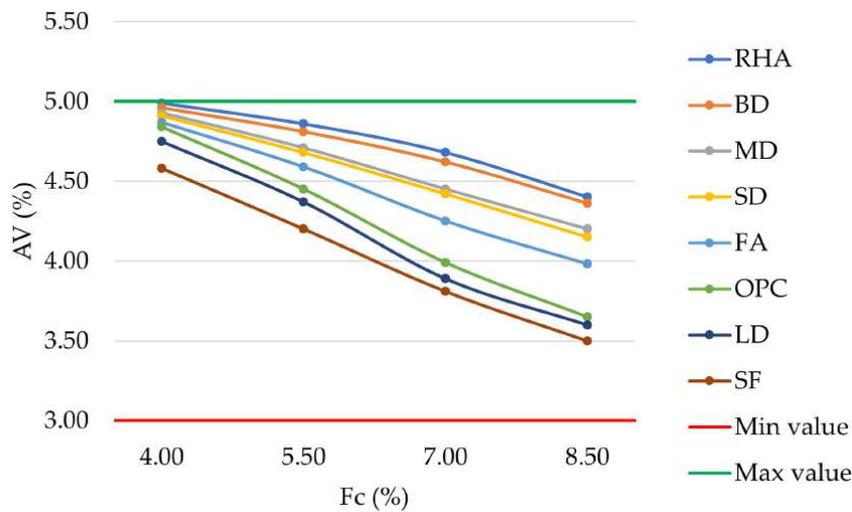


Figura 8.4: AV in funzione di Fc per conglomerati con filler convenzionali e di riciclo.

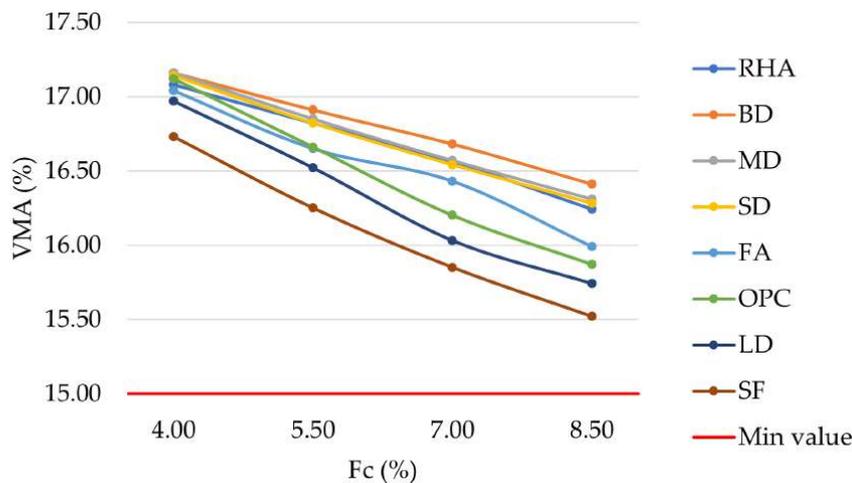


Figura 8.5: VMA in funzione di Fc per conglomerati con filler convenzionali e di riciclo.

e RHA hanno restituito valori di VFB inferiori rispetto a quelli mostrati dalle miscele OPC, rispettivamente pari al 75,10%, 74,49%, 74,27%, 73,45% e 72,80%. Grazie alla ridotta possibilità di risalita del bitume, le miscele preparate con questi cinque filler alternativi potrebbero essere preferite nelle regioni a clima caldo [Kutuk-Sert, 2016].

Il comportamento meccanico delle miscele analizzate è stato descritto empiricamente mediante i parametri di Marshall. MS è rappresentata graficamente nella Figura 8.7. Come da aspettative, un aumento della percentuale di filler ha provocato un aumento della MS per ciascuna percentuale di bitume [Nikolaides, 2015, Akbulut et al., 2012, Huang et al., 2007]. Ad

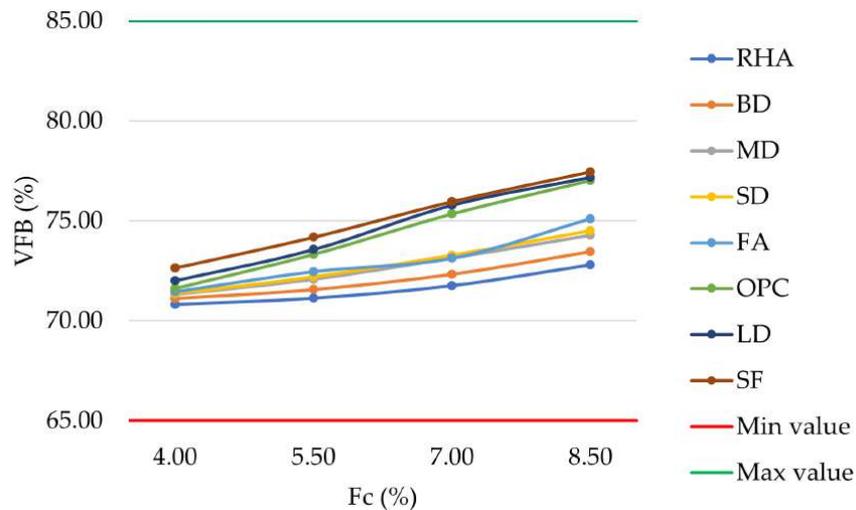


Figura 8.6: VFB in funzione di Fc per conglomerati con filler convenzionali e di riciclo.

eccezione delle miscele preparate con il 4%, 5,5% e 7% di RHA e delle miscele preparate con il 4% di MD (9,01 kN, 9,36 kN, 9,80 kN e 9,81 kN, rispettivamente), tutte le altre miscele hanno restituito valori di MS superiori alla soglia di accettazione di 10 kN stabilita dalla normativa indiana. Il valore più alto di MS più alto è stato raggiunto per il contenuto di filler più elevato, con valori rispettivamente pari a 10,18 kN, 12,31 kN, 12,39 kN, 12,60 kN, 12,93 kN, 13,74 kN, 13,86 kN e 13,98 kN per le miscele preparate con RHA, MD, FA, BD, SD, OPC, LD e SF. Le miscele LD e SF hanno mostrato un comportamento simile a quello dell'OPC, dimostrando che entrambi i materiali sono in grado di sostituire efficacemente l'OPC nelle miscele bituminose. Infine, per contenuti di filler pari al 5,5%, 7,0% e 8,5%, le miscele preparate con questi tre filler hanno soddisfatto anche il requisito di accettazione di 13 kN richiesto dagli standard MoRTH per le miscele preparate con bitume modificato.

L'indagine di laboratorio relativa ai valori di MF ha prodotto risultati contenuti nell'intervallo 2 mm÷4 mm, come richiesto dagli standard MoRTH. Come previsto, i valori più bassi di MF si sono verificati in corrispondenza dei più alti contenuti di filler. Identificando la maggiore stabilità e i valori di scorrimento più bassi, è possibile determinare empiricamente le miscele più resistenti in termini di deformazioni permanenti alle alte temperature. Per ciascuna miscela analizzata, l'andamento del MQ può essere osservato in Figura 8.8.

In generale, l'aumento del contenuto di filler ha condotto a valori più elevati di MQ. La sostituzione dell'OPC con LD o SF ha aumentato leggermente il valore di MQ nelle miscele corrispondenti, con differenze che vanno dallo 0,04% allo 0,37%. Tuttavia, i risultati hanno confermato il raggiungimento dei requisiti di accettazione previsti dal MoRTH (2 kN/mm÷6 kN/mm) per tutte le miscele analizzate. Per quanto riguarda le miscele SD, MD, FA, BD e RHA, i valori di MQ sono risultati inferiori a quelli ottenuti con l'OPC.

Per comprendere il comportamento meccanico in termini di resistenza a trazione indiretta dei conglomerati bituminosi analizzati, in Figura 8.9 è possibile apprezzare l'andamento dell'ITS al

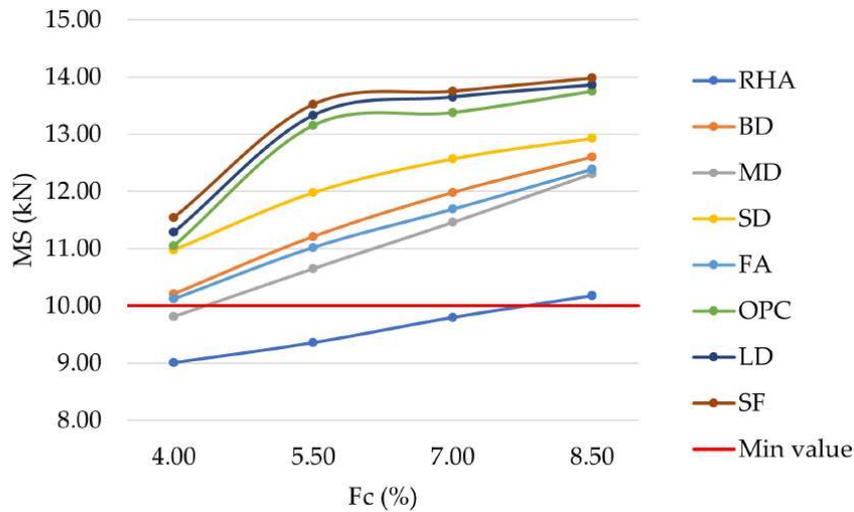


Figura 8.7: MS in funzione di Fc per conglomerati con filler convenzionali e di riciclo.

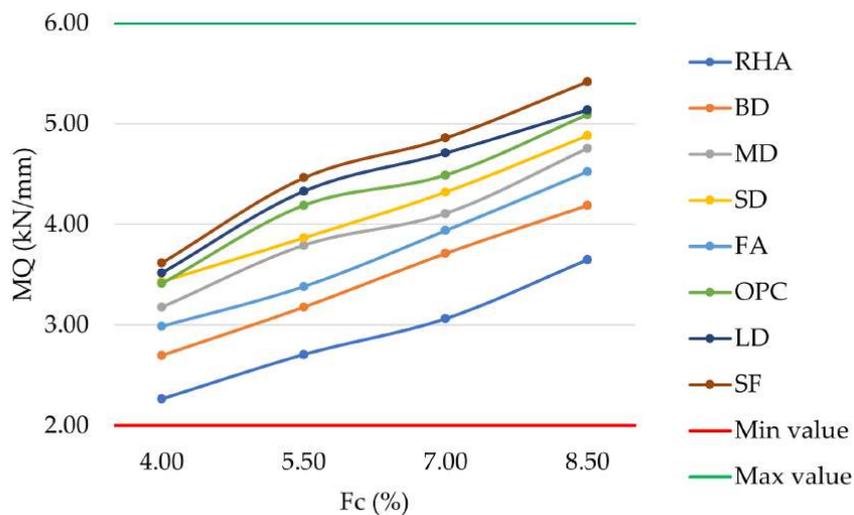


Figura 8.8: MQ in funzione di Fc per conglomerati con filler convenzionali e di riciclo.

variare del contenuto di filler. Come si può notare, un aumento del contenuto di filler ha condotto ad un corrispondente aumento dell'ITS, con i valori più elevati riscontrati in corrispondenza di un contenuto di filler pari all'8,5%, indicativi di una maggiore resistenza alla fessurazione per fatica. La sostituzione dell'OPC con LD o SF ha lasciato pressoché inalterata la risposta in termini di ITS, aumentandone leggermente i valori. Le differenze si sono attestate tra 1 kPa e 11 kPa, dimostrando che i filler LD e SF siano in grado di sostituire adeguatamente l'OPC nei conglomerati bituminosi. Gli standard MoRTH richiedono un ITS superiore a 600 kPa

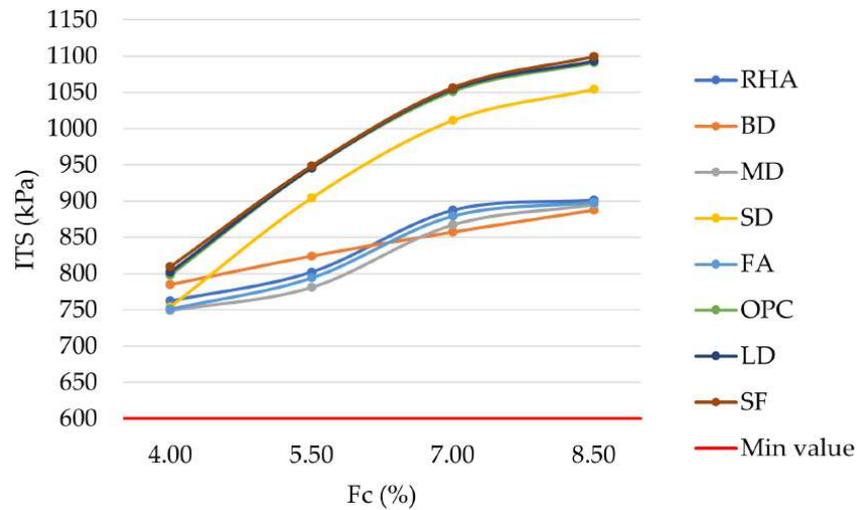


Figura 8.9: ITS in funzione di Fc per conglomerati con filler convenzionali e di riciclo.

[Ministry of Road Transport and Highways, 2013]: tale condizione è stata verificata per tutte le miscele esaminate, con valori ITS di picco anche superiori ad 1 MPa.

L'indagine di laboratorio sull'ITSR (Figura 8.10) ha prodotto risultati superiori al 75%, così come previsto dagli standard MoRTH. Un contenuto di filler più elevato ha determinato una resistenza leggermente superiore ai danni provocati dal condizionamento in acqua, con le miscele LD e SF che hanno raggiunto valori di ITSR più elevati rispetto a quelli raggiunti dalle miscele OPC. Le differenze sono state comprese tra lo 0,19% e l'1,30%.

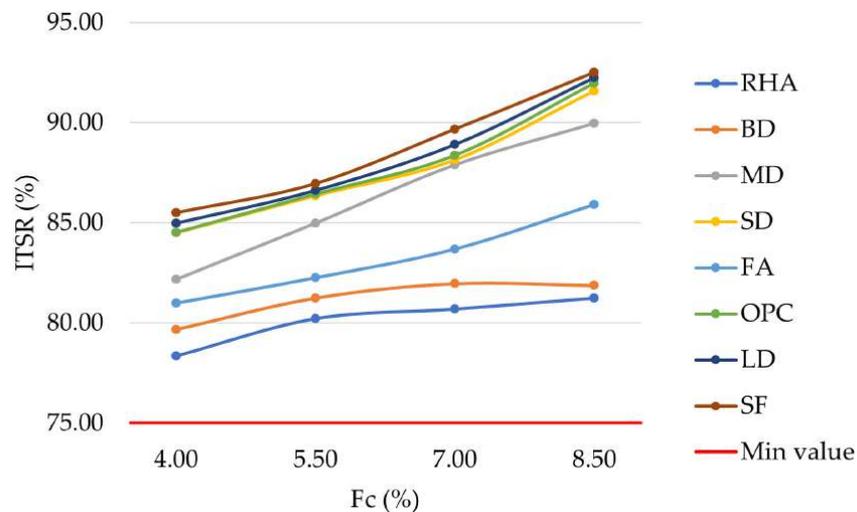


Figura 8.10: ITSR in funzione di Fc per conglomerati con filler convenzionali e di riciclo.

Grazie alle proprietà attivanti dell'adesione della portlandite e della calcite presenti nel filler SF, le corrispondenti miscele hanno fatto registrare i valori di ITSR più elevati. Analogamente, grazie alla calcite presente in LD, l'adesione del bitume è stata migliorata con conseguenti elevati valori di ITSR. Le miscele SD hanno evidenziato una resistenza al danno derivante dall'acqua simile a quella delle miscele OPC; mentre MD, FA, BD e RHA hanno riportato valori di ITSR inferiori. Le argille fini contenute in BD hanno ridotto la resistenza al danno da condizionamento in acqua delle miscele corrispondenti. Risultati simili sono stati osservati anche da [Arabani et al., 2017, Kuity et al., 2014]. Infine, l'elevata porosità dell'RHA ha determinato i più bassi valori di ITSR osservati nell'ambito di questa campagna sperimentale.

Per quanto riguarda la resistenza all'abrasione, si può osservare che contenuti di filler più elevati si sono sempre tradotti in valori di CL più bassi (Figura 8.11). Le miscele LD e SF hanno fatto registrare risultati migliori rispetto a quelle preparate con OPC convenzionale. Considerando un contenuto di filler pari all'8,5%, i valori di CL ottenuti per le miscele preparate con RHA, BD, FA, MD, SD, OPC, LD e SF sono stati rispettivamente pari al 20,25%, 19,01%, 15,15%, 10,89%, 8,97%, 8,52%, 8,42% e 8,04%. In accordo con la Figura 8.4, i valori più bassi dei vuoti d'aria hanno comportato una minore insorgenza di fenomeni di *raveling* (o disgregazione degli aggregati).

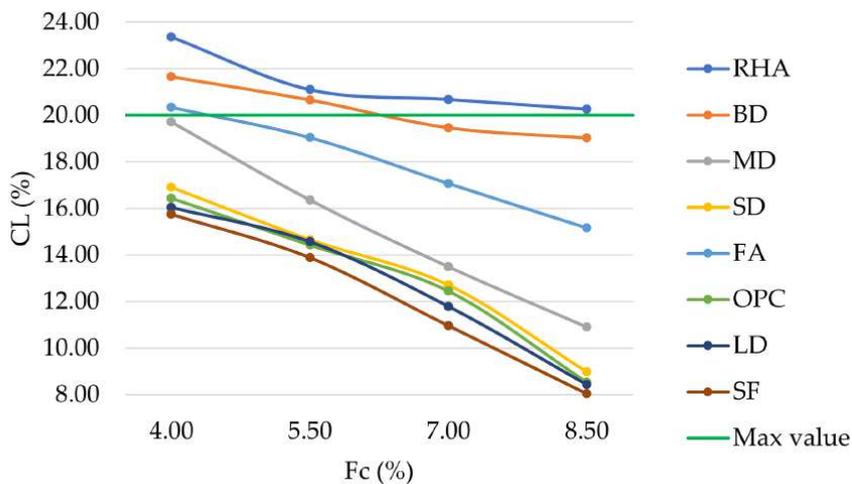


Figura 8.11: CL in funzione di Fc per conglomerati con filler convenzionali e di riciclo.

8.4.2 Risultati della modellazione predittiva

La descrizione statistica delle variabili coinvolte nella modellazione predittiva è stata riportata in Tabella 8.5. Come precedentemente illustrato, sono state analizzate otto tipologie e quattro diverse percentuali di filler, avendo fissato sia il tipo e il contenuto di bitume, sia la granulometria degli aggregati. Questo ha prodotto 32 osservazioni originali per ciascuna variabile che, combinate con i 24 dati sintetici generati mediante l'interpolazione Makima, hanno consentito di generare una banca dati completa composta da 56 osservazioni per ciascuna specifica variabile.

Tabella 8.5: Descrizione statistica delle variabili coinvolte nel modello predittivo.

Variabile	Descrizione	Unità di misura	Numero di osservazioni	Media	Dev. Std
CV	Variabile categorica (tipologia di filler)	-	56	-	-
Fc	Contenuto di filler	%	56	6.25	1.51
AV	Vuoti d'aria	%	56	4.42	0.40
MS	Stabilità Marshall	kN	56	11.88	1.42
MQ	Quoziente Marshall	kN/mm	56	3.93	0.74
ITS	Resistenza a trazione indiretta	kPa	56	901.61	107.82
ITSR	Rapporto di resistenza a trazione indiretta	%	56	85.35	3.76
CL	Resistenza all'abrasione	%	56	15.73	4.05

Una matrice delle correlazioni di Pearson [Pallant, 2020] è stata utilizzata per determinare in via preliminare il grado di correlazione esistente tra le variabili considerate dal modello predittivo (Figura 8.12). Ogni elemento della matrice è descritto da un segno e da un valore assoluto. Il primo è indice di una proporzionalità diretta (+) oppure inversa (-) tra le due variabili considerate; il secondo, invece, riporta in modo specifico l'intensità della correlazione e varia tra 0 e 1. Un valore pari a 0 significa che non c'è alcuna correlazione, ovvero il valore di una variabile non può essere determinato a partire dal valore dell'altra. Al contrario, un valore pari a 1 significa che esiste una correlazione perfetta, ovvero il valore di una variabile può essere determinato esattamente in base al valore dell'altra. A titolo esemplificativo, è stata individuata una forte correlazione positiva tra CV e MS ($R=+0.76$) e tra Fc e ITS ($R=+0.70$). Viceversa, è stata identificata una forte correlazione negativa tra AV e MQ ($R=-0.92$).

Per migliorare l'efficienza dell'addestramento del modello [Saleh and Fleyeh, 2022], tutte le osservazioni sperimentali di ciascuna variabile sono state normalizzate:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (8.4.1)$$

Tutti i dati sono stati pertanto scalati al medesimo intervallo $[0, +1]$, in modo che i valori minimi e massimi di ciascuna variabile corrispondessero rispettivamente ai limiti inferiore e superiore di tale intervallo.

Il processo di addestramento del modello CatBoost è stato rappresentato in Figura 8.13. Inizialmente, sia la funzione di perdita in fase di addestramento che quella in fase di validazione hanno restituito valori piuttosto elevati, pari a circa 0,60. Con il procedere delle iterazioni, si è osservata una rapida diminuzione di entrambe le funzioni di perdita, con valori che si sono assestati intorno a 0,20 dopo circa 50 e 100 iterazioni, rispettivamente per la funzione di perdita in fase di addestramento e per quella in fase di validazione. Il valore di MultiRMSE è diminuito gradualmente fino al raggiungimento del punto ottimale, identificato all'iterazione numero 268 da un valore di funzione di perdita in fase di validazione pari a 0,1569. Superato tale punto, non

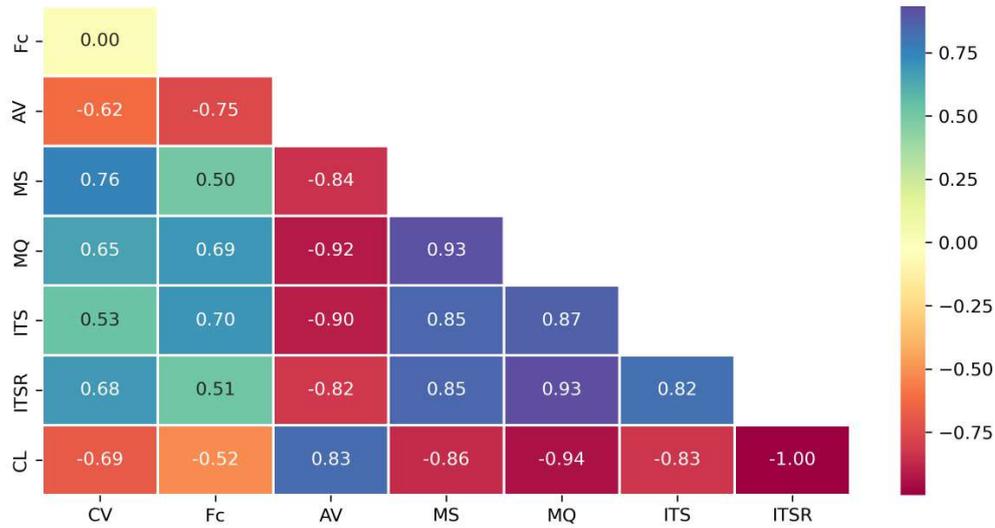


Figura 8.12: Matrice delle correlazioni di Pearson.

è più apprezzabile una diminuzione significativa del valore di MultiRMSE in fase di validazione. Pertanto, il processo di addestramento è stato interrotto dal rilevatore di *overfitting* al termine di 20 ulteriori iterazioni. L'identificazione del punto ottimale ha consentito di identificare e fissare la migliore configurazione degli iperparametri del modello e di procedere alla successiva fase di test.

In seguito alla denormalizzazione delle variabili elaborate, è iniziata la fase di test. L'affidabilità delle previsioni del modello è stata descritta attraverso sei diverse metriche di valutazione, ovvero MAE, MAPE, MSE, RMSE, R e R^2 (Tabella 8.6). Per ognuna delle cinque variabili target, l'errore percentuale medio assoluto è sempre risultato inferiore al 5%, con coefficienti di correlazione di Pearson superiori a 0,95 e coefficienti di determinazione superiori a 0,88.

Tabella 8.6: Metriche di valutazione delle prestazioni del modello CatBoost.

Metrica	MS	MQ	ITS	ITSR	CL
MAE	0.2595	0.1099	29.5495	0.6441	0.7317
MAPE	2.3328	3.0829	3.4862	0.7669	4.3340
MSE	0.1101	0.0195	1432.5907	0.6495	0.7670
RMSE	0.3319	0.1396	37.8496	0.8059	0.8758
R	0.9778	0.9826	0.9543	0.9744	0.9727
R^2	0.9478	0.9626	0.8885	0.9446	0.9437

L'influenza di ciascuna variabile sulle previsioni effettuate dal modello è stata determinata mediante un'analisi di sensitività implementata in Python 3.8.5 (Figura 8.14). L'algoritmo consente di determinare l'importanza di ciascuna variabile di input e di normalizzarla affinché la somma dell'importanza di ciascun input raggiunga un totale del 100%. Quanto più importante è una determinata variabile, tanto più significativa sarà la variazione media delle previsioni se

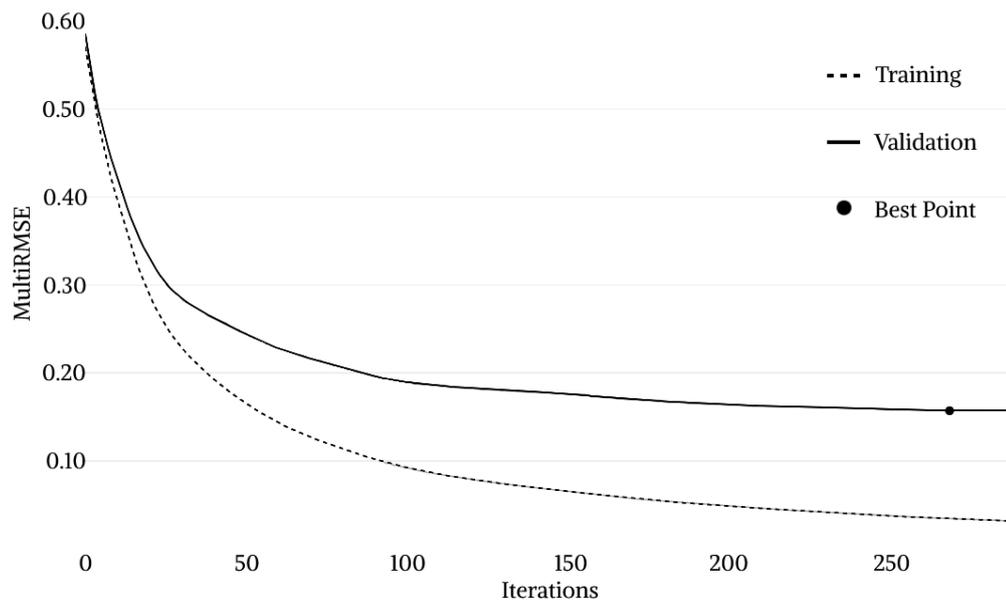


Figura 8.13: Processo di addestramento del modello CatBoost.

tale variabile dovesse essere modificata. Si può notare come la variabile categorica possieda la maggiore importanza (47,89%), seguita dal contenuto di vuoti d'aria (39,07%) e dal contenuto di filler (13,04%).

Al fine di fornire una visione più approfondita delle prestazioni del modello, nella Figura 8.15 sono stati riportati i confronti diretti tra le osservazioni dei vettori di test e le corrispondenti predizioni effettuate dal modello CatBoost. Le prime sono rappresentate da istogrammi neri mentre le seconde da istogrammi grigi. L'ID identificativo di ciascuna osservazione di test è stato indicato sull'asse orizzontale. Sebbene si possano osservare fluttuazioni rilevanti nei valori di ciascuna variabile obiettivo, le differenze tra gli istogrammi corrispondenti ai valori reali e quelli corrispondenti ai valori previsti non si sono mai rivelate particolarmente pronunciate.

Un'ulteriore prospettiva sull'accuratezza previsionale del modello CatBoost è stata fornita dai grafici di regressione rappresentati in Figura 8.16. Il nome della variabile corrispondente può essere visualizzato nella parte inferiore di ciascun grafico, mentre i rispettivi coefficienti di correlazione di Pearson sono stati riportati nella parte superiore. I valori stimati mediante modello predittivo sono stati rappresentati da cerchi azzurri, mentre la perfetta corrispondenza teorica tra valori veri e quelli predetti (corrispondente ad una correlazione $R=1$) è stata rappresentata dalla linea di uguaglianza riportata in blu. La distanza tra valori predetti e linea di uguaglianza è risultata sempre ridotta, per ciascuna osservazione di test di ciascuna variabile obiettivo, ad ulteriore conferma dei notevoli risultati ottenuti in riferimento ai coefficienti di correlazione di Pearson.

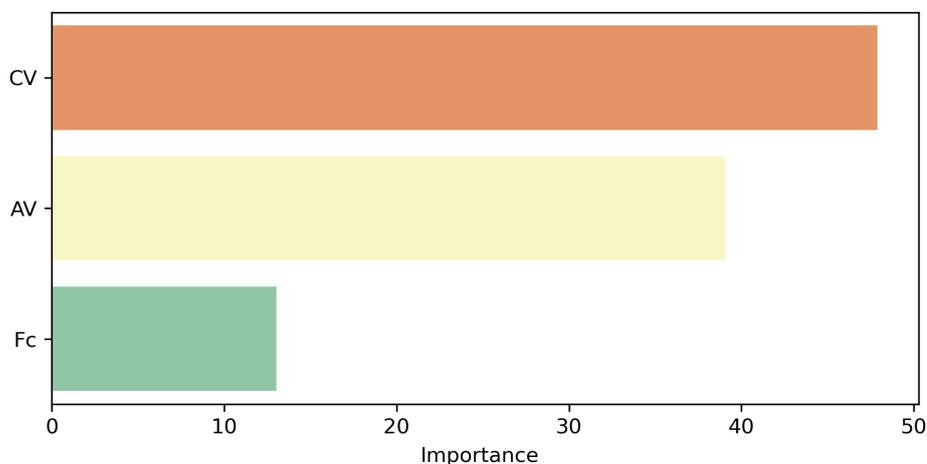


Figura 8.14: Risultati dell'analisi di sensitività.

8.5 Conclusioni

Lo studio presentato in questo capitolo descrive un'indagine di laboratorio eseguita su sette selezionati filler alternativi che potrebbero potenzialmente sostituire l'OPC convenzionalmente utilizzato nei conglomerati bituminosi; successivamente, un modello di apprendimento automatico realizzato mediante un approccio CatBoost basato su alberi decisionali è stato sviluppato ed implementato per la previsione simultanea delle prestazioni meccaniche di tali miscele. Tutti i parametri di progetto delle miscele sono stati mantenuti costanti, ad eccezione del tipo e della percentuale di filler, scelto tra RHA, BD, MD, SD, FA, LD e SF e variabile tra il 4,0% e l'8,5% con un incremento dell'1,5%. Per un'accurata valutazione sperimentale sono stati eseguiti test Marshall, quelli di resistenza alla trazione indiretta, di sensibilità al condizionamento in acqua e di resistenza all'abrasione. I risultati ottenuti sono stati successivamente implementati in una metodologia ML che ha consentito di prevedere simultaneamente MS, MQ, ITS, ITSR e CL sulla base di tre input, ovvero: una variabile categorica identificativa della tipologia di filler, l'effettivo contenuto percentuale del filler stesso e la percentuale di vuoti d'aria. Per valutare nel dettaglio il grado di attendibilità delle previsioni, sono state implementate sei diverse metriche di bontà di adattamento: MAE, MAPE, MSE, RMSE, R e R^2 . Sono stati ottenuti risultati incoraggianti, con valori di MAPE sempre inferiori al 5% e valori di R e R^2 sempre rispettivamente superiori a 0,95 e 0,88 per ciascuna delle cinque variabili obiettivo. Sulla base dei risultati ottenuti, è possibile trarre le seguenti conclusioni:

- Sulla base della caratterizzazione chimico-fisica, tutti i filler alternativi analizzati potrebbero essere potenzialmente utilizzati nella progettazione di conglomerati bituminosi in sostituzione dell'OPC secondo gli standard MoRTH;
- I requisiti MoRTH sono stati pienamente soddisfatti anche in termini di resistenza meccanica. Ad eccezione delle miscele preparate con 4,0%, 5,5% e 7,0% di RHA e di quelle

preparate con 4,0% di MD, tutti gli altri conglomerati hanno mostrato un valore di MS superiore alla soglia di accettazione di 10 kN. LD e SF all'8,5% hanno addirittura fatto registrare un valore di MS superiore al requisito di accettazione per le miscele con bitume modificato, pari a 13 kN;

- Tutte le miscele analizzate hanno soddisfatto anche le prescrizioni del MoRTH in termini di sensibilità all'acqua, poiché hanno restituito valori di ITSR costantemente superiori al 75%;
- In generale, tra tutte le possibilità analizzate, LD e SF sono risultati i migliori filler alternativi per le miscele bituminose. Non solo hanno soddisfatto gli standard MoRTH, ma hanno anche fatto registrare prestazioni migliori rispetto a quelle derivanti dall'utilizzo del convenzionale OPC in termini di MS, ITS, ITSR, e CL;
- La tecnica di *data augmentation* si è dimostrata efficace. Sono stati ottenuti risultati notevoli in termini di tutte le metriche di valutazione delle prestazioni del modello predittivo, anche sulla base di una banca dati iniziale caratterizzata da dimensioni ridotte;
- L'approccio CatBoost ha consentito di ottenere uno strumento predittivo affidabile, in grado di fornire previsioni attendibili delle prestazioni meccaniche delle miscele, evitando così ulteriori procedure sperimentali costose ed onerose in termini di tempo;
- L'intera metodologia è stata sviluppata utilizzando Python 3.8.5 e risulta, pertanto, facilmente interpretabile e implementabile anche da altri ricercatori. Tuttavia, è opportuno ricordare che qualsiasi ulteriore applicazione a banche dati diverse da quella analizzata nel presente capitolo richiede una nuova calibrazione e successiva ottimizzazione degli iperparametri del modello.

L'intera indagine sperimentale è stata eseguita su scala di laboratorio. Un importante sviluppo futuro potrebbe essere rappresentato dall'incremento delle dimensioni della banca dati disponibile, ottenuto ad esempio variando il contenuto di bitume delle miscele o altri parametri progettuali. Inoltre, potrebbe essere interessante includere altre proprietà legate alle prestazioni fisico-meccaniche delle miscele, come ad esempio la resistenza all'ormaiamento oppure la vita a fatica. Infine, sarebbe curioso eseguire una simile indagine su scala reale per comprendere l'effettivo comportamento meccanico mostrato sul campo dalle miscele preparate con i riempitivi alternativi studiati e confrontare i risultati ottenuti con quelli che si possono evincere dalla ricerca descritta.

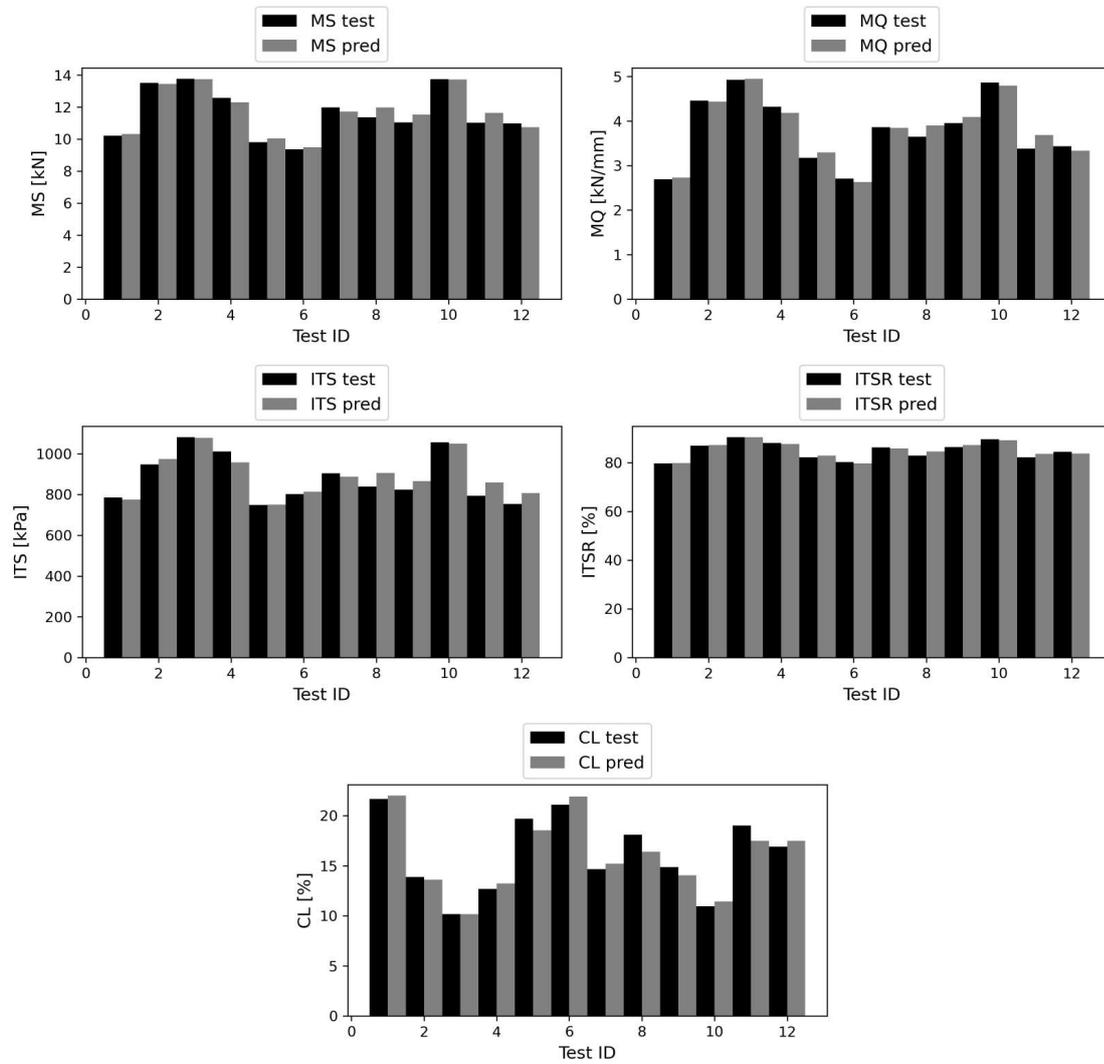


Figura 8.15: Confronti tra test e predizioni CatBoost per ciascuna variabile obiettivo.

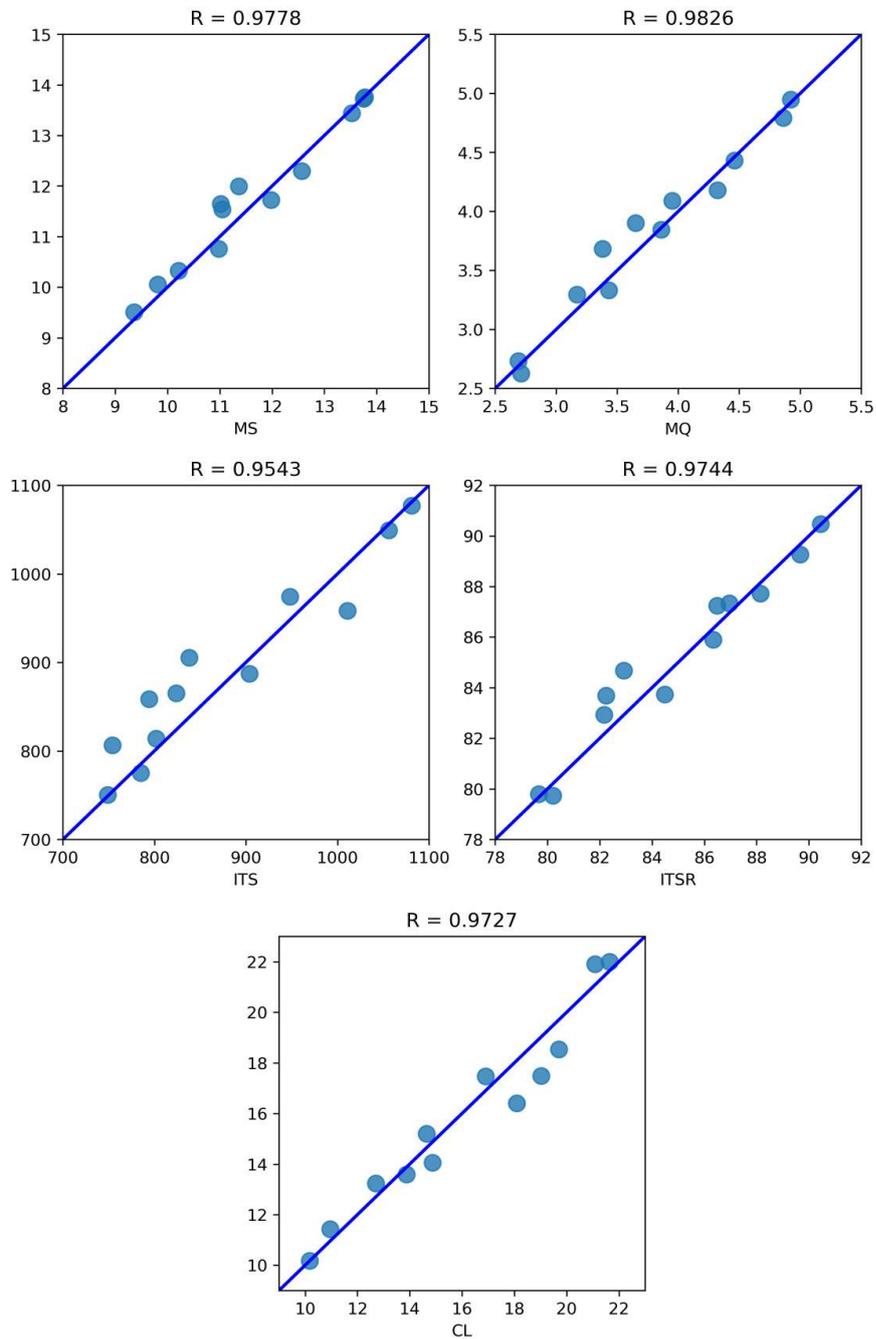


Figura 8.16: Grafici di regressione per ciascuna variabile obiettivo.

Capitolo 9

Modulo Complesso di Conglomerati Bituminosi

” You can have data without information, but you cannot have information without data.

– Daniel Keys Moran

9.1 Introduzione

Una delle principali caratteristiche tecniche atte ad analizzare il comportamento meccanico dei conglomerati bituminosi e valutarne le rispettive performance è rappresentata dal modulo complesso E^* . Questo fondamentale parametro consente di descrivere accuratamente la variazione della rigidità di una miscela al variare delle condizioni di prova, ovvero della frequenza di carico e della temperatura di prova [Ali et al., 2015]. Il modulo complesso è definito da due componenti principali: il modulo dinamico $|E^*|$ e l'angolo di fase φ . Il primo descrive la rigidità del materiale a determinate condizioni di temperatura e frequenza, mentre il secondo determina le corrispondenti componenti elastiche e viscosive alle stesse condizioni di prova. Mediante il principio di sovrapposizione tempo-temperatura (*Time-Temperature Superposition Principle* - TTSP) [Chehab et al., 2002], questi dati vengono solitamente elaborati per ottenere le cosiddette curve maestre. Tali curve continue consentono di stimare il modulo complesso della miscela in esame per un'ampia gamma di combinazioni di temperatura e frequenza. In linea generale, il modulo complesso è un parametro cruciale sia per la progettazione di pavimentazioni in conglomerato bituminoso sia per la valutazione delle loro performance meccaniche [Han et al., 2021] ed è richiesto come input dalle principali procedure di progettazione come la *Mechanistic-Empirical Pavement Design Guide* (M-EPDG) [AASHTO, 1993].

Per determinare il modulo complesso e l'angolo di fase di un conglomerato bituminoso si utilizzano principalmente due metodi:

1. è possibile eseguire prove di laboratorio su diversi campioni della miscela in esame;

2. in alternativa, si possono utilizzare equazioni empiriche predittive per stimare i suddetti parametri.

Tuttavia, per la determinazione sperimentale di $|E^*|$ e φ è prevista una procedura molto onerosa in termini di tempo: bisogna, infatti, preparare diversi campioni compattati e testarli per ampi intervalli sia di temperatura che di frequenza per mezzo di adeguate attrezzature utilizzate da tecnici altamente qualificati [Ali et al., 2017]. Pertanto, gli sforzi di molti ricercatori si sono quindi orientati verso i metodi predittivi, limitando l'impiego delle onerose prove di laboratorio e prevedendo $|E^*|$ e φ per mezzo di equazioni semi-empiriche sviluppate ad hoc. I primi tentativi di successo hanno condotto all'equazione predittiva introdotta da Fonseca e Witczak [Fonseca and Witczak, 1996], seguita dal consolidato modello di Witczak basato sulla viscosità [Bari and Witczak, 2006]. In input sono richiesti elementi quali la granulometria degli aggregati, la percentuale di vuoti d'aria (*Air Voids* - AV), il contenuto effettivo di legante bituminoso ($V_{b_{eff}}$), nonché la viscosità del bitume η_b e la frequenza di carico (f_i). In output, invece, viene restituito il valore del modulo dinamico. Le previsioni si sono dimostrate affidabili ed accurate: tuttavia, poiché entrambe le metodologie sono state sviluppate a partire da osservazioni relative a miscele standard, queste hanno dimostrato di possedere alcuni limiti di applicabilità [Zhang et al., 2022].

Molti altri modelli sono stati sviluppati ed utilizzati più o meno frequentemente per produrre previsioni di modulo dinamico dei conglomerati bituminosi [Olard and Di Benedetto, 2003, Di Benedetto et al., 2004, Witczak et al., 2007, Li et al., 2012]. Con il passare degli anni, seppur in numero più limitato, sono stati proposti anche modelli regressivi per la previsione dell'angolo di fase [Biligiri et al., 2010, Nemati and Dave, 2018]. Tuttavia, per comprendere meglio le relazioni che sussistono tra il modulo complesso e le diverse variabili che lo influenzano, i modelli regressivi sono stati progressivamente sostituiti dagli algoritmi di apprendimento automatico (ML) [Ceylan et al., 2008, Komba et al., 2012, Sakhaeifar et al., 2015, Cao et al., 2016, Liu et al., 2017]. A differenza dei modelli basati sulla regressione, questi algoritmi innovativi hanno la straordinaria capacità di modellare e prevedere fenomeni anche molto diversi tra loro, le cui relazioni fondamentali non devono essere necessariamente conosciute a priori. Tra le più diffuse e praticate tecniche di ML è possibile identificare le reti neurali artificiali che hanno dimostrato di essere in grado di comprendere anche fenomeni altamente non lineari e di produrre previsioni particolarmente accurate [Svilar et al., 2022, Althoey et al., 2023, Baldo et al., 2021, Baldo et al., 2022b]. Tuttavia, presentano alcuni inconvenienti come la complessità nell'identificare la topologia e gli iperparametri migliori così come la scarsa interpretabilità del modello complessivo.

Un'altra tipologia di algoritmo di apprendimento automatico analogamente adatto a risolvere problemi di regressione si basa sugli alberi decisionali ed è in grado di restituire previsioni altrettanto accurate in applicazioni anche molto differenti tra loro [Karbassi et al., 2014, Poh et al., 2018, Upadhyay et al., 2023]. Tali approcci consistono nel generare un modello predittivo ad albero, la cui struttura si determina mediante semplici regole decisionali che ne determinano una più semplice interpretabilità [Baduge et al., 2022]. Alcune applicazioni per la risoluzione di problemi di regressione legati al modulo dinamico sono descritte e possono essere approfondite in modo esaustivo nella letteratura scientifica di riferimento [Daneshvar and Behnood, 2022, Ali et al., 2021]. L'obiettivo di questo capitolo è quello di sviluppare un modello di apprendimento automatico avanzato, in grado di implementare un algoritmo innovativo basato sugli alberi decisionali, il *Categorical Boosting* o *CatBoost*, al fine di prevedere accuratamente sia

il modulo dinamico che l'angolo di fase di molteplici miscele di conglomerato bituminoso. La tecnica CatBoost è il risultato della combinazione di algoritmi innovativi di elaborazione delle informazioni incentrati sull'efficienza computazionale e sull'elevata accuratezza predittiva, fattori che l'hanno resa superiore ad altri principali algoritmi concorrenti di *gradient-boosting* [Bentéjac et al., 2021]. Un'ampia campagna sperimentale di prove di flessione su quattro punti (4PBT) è stata condotta e sono state registrate 1680 osservazioni: le informazioni ottenute sono state poi utilizzate per addestrare il modello al fine di prevedere simultaneamente sia il modulo dinamico che l'angolo di fase di diversi conglomerati bituminosi. Al fine di confrontare in maniera equa il modello sviluppato con lo stato dell'arte (State-of-Art - SoA) delle reti neurali artificiali e con due equazioni empiriche ben consolidate, le caratteristiche di input assegnate sono rimaste invariate in tutti e quattro i diversi approcci. Inoltre, è stata eseguita un'approfondita analisi di sensitività per comprendere al meglio il funzionamento del modello e l'impatto che le diverse caratteristiche di input hanno rispettivamente avuto sugli output predetti. Considerando le notevoli prestazioni raggiunte dal modello, questo potrebbe fornire uno strumento affidabile per la previsione simultanea dei valori di $|E^*|$ e φ da implementare in un secondo momento nelle più comuni procedure di progettazione delle miscele bituminose. Il modello messo a punto conferma quanto già noto sulle correlazioni tra variabili compositive e parametri prestazionali e potrebbe portare ad una comprensione ancora più approfondita delle relazioni esistenti tra queste variabili ed il corrispondente comportamento meccanico dei conglomerati bituminosi. In questo senso, la ricerca qui descritta può rappresentare un contributo alla letteratura scientifica esistente ed un interessante spunto per ulteriori applicazioni future.

9.2 Campagna Sperimentale

9.2.1 Selezione dei materiali e test di modulo dinamico

Al fine di costituire un'ampia banca dati utile per le successive attività di modellazione predittiva, sono stati progettati e prodotti nove diversi conglomerati bituminosi: questi sono stati dunque testati per ottenere informazioni sulle corrispondenti risposte viscoelastiche. In Figura 9.1, è possibile osservare la rappresentazione schematica dell'intera impostazione metodologica adottata.

Il diabase è stato utilizzato sia come aggregato grossolano che fine durante la preparazione delle miscele per strati superficiali, indicate nel seguito con le sigle AC11 e SMA11 (*Asphalt Concrete* e *Stone Mastic Asphalt*, 11 rappresenta la dimensione nominale massima degli aggregati espressa in mm). Nelle miscele per strati di binder e base (AC22 e AC32), invece, la scelta per la natura dell'aggregato è ricaduta sul calcare tradizionale. In tutti i casi, il filler è stato costituito da polvere di calcare. Sono stati utilizzati sia leganti convenzionali (non modificati) che modificati per mezzo di polimeri. La granulometria degli aggregati è conforme ai corrispondenti standard europei per conglomerati bituminosi [UNI EN 13108-1, 2016]. Il contenuto ottimale di legante bituminoso è stato determinato tramite procedura Marshall. Maggiori dettagli circa le composizioni delle diverse miscele e le curve granulometriche sono stati riportati rispettivamente in Tabella 9.1 e Figura 9.2.

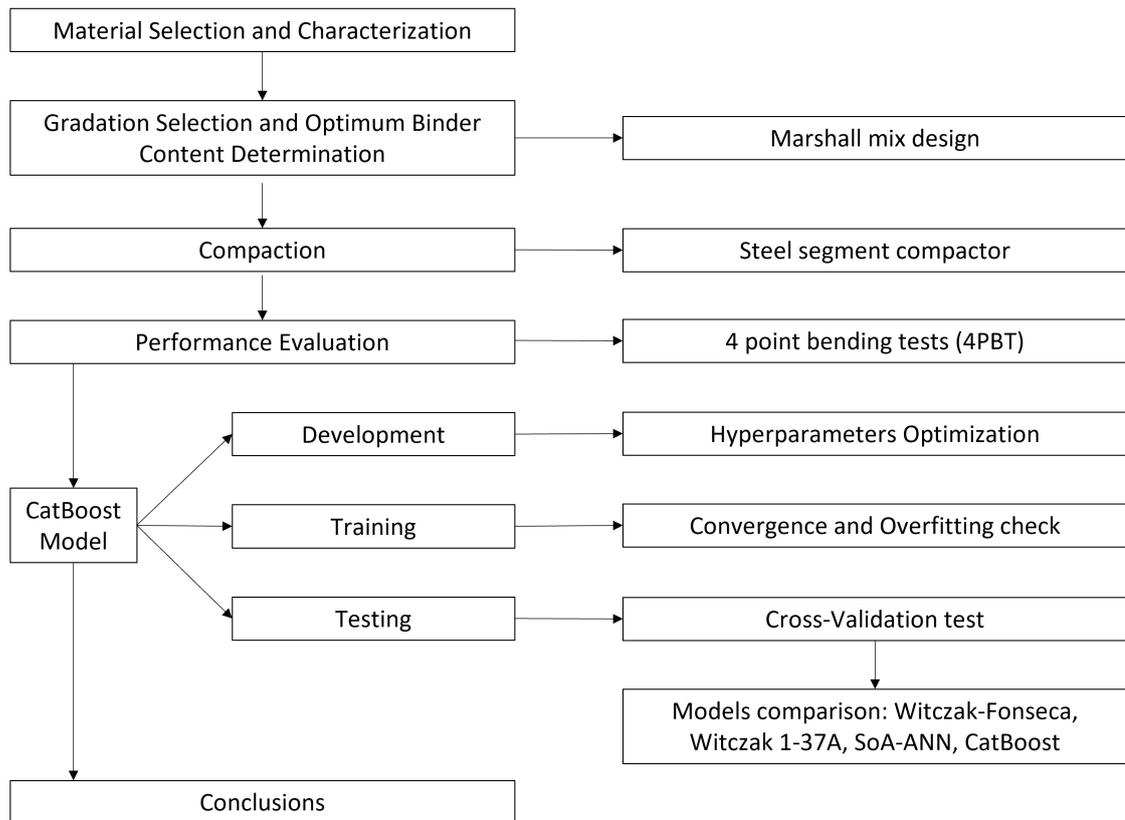


Figura 9.1: Rappresentazione schematica della metodologia seguita.

Per tutte le miscele, le lastre di conglomerato bituminoso sono state compattate tramite *Steel Segment Compactor* secondo quanto prescritto dalla normativa [UNI EN 12697-33, 2019]. Da ciascuna lastra sono stati poi ottenuti tre campioni di dimensioni nominali pari a 60x60x500 mm. Una volta determinate le proprietà volumetriche di ciascun provino, i valori di rigidità sono stati determinati attraverso prove di flessione su 4 punti (*4 Point Bending Test*) in controllo di temperatura e frequenza di carico. Per ciascuna miscela, tre diversi campioni sono stati testati. I risultati di modulo dinamico ed angolo di fase sono stati quindi ottenuti tramite le equazioni riportate nella normativa di riferimento [UNI EN 12697-26, 2018] e calcolando i valori medi delle prove eseguite sui singoli campioni.

9.2.2 Equazioni empiriche per $|E^*|$

Nel corso degli anni, diversi modelli sono stati studiati e sviluppati per determinare una relazione tra il modulo dinamico $|E^*|$ ed i principali parametri fisici dei conglomerati bituminosi. Uno tra i più validi e scientificamente apprezzati è stato proposto da Witczak nel 1996 e si basa sul principio TTSP. I valori di E^* determinati alle diverse temperature di test vengono traslati orizzontalmente e riferiti ad un'unica temperatura di riferimento al fine di ottenere la cosiddetta

Tabella 9.1: Descrizione riassuntiva dei conglomerati bituminosi analizzati.

Miscela	Aggregato	Bitume	AV (%)	Stabilità (kN)	Densità rel.
AC11_70_100	Diabase	5.6% PG63-24	3.4%	12.6	2.56
AC11_PmB45_80_50	Diabase	5.6% PG69-27	3.7%	13.9	2.56
AC11_PmB45_80_65	Diabase	5.6% PG76-25	3.4%	13.3	2.56
AC22_mB160_220FT	Calcare	4.5% PG73-22	3.8%	11.1	2.62
AC22_PmB45_80_65	Calcare	4.5% PG76-25	4.1%	13.3	2.61
AC22_50_70	Calcare	4.5% PG67-24	4.4%	11.2	2.62
AC32_50_70	Calcare	4.3% PG67-24	3.2%	–	2.64
SMA11_70_100	Diabase	6.5% PG63-24	3.6%	7.2	2.54
SMA11_PmB45_80_65	Diabase	6.5% PG76-25	3.4%	9.9	2.52

curva maestra [Fonseca and Witzak, 1996]. Per i conglomerati bituminosi, tale curva viene assunta come sigmoidale non lineare dal momento che si ritiene sia in grado di rappresentare correttamente la dipendenza del comportamento meccanico della miscela dalla temperatura:

$$\log|E^*| = \rho + \frac{\alpha}{1 + e^{\beta + \gamma \log(f_{red})}} \quad (9.2.1)$$

$|E^*|$ rappresenta il modulo dinamico, ρ il logaritmo del valore minimo di $|E^*|$, $\rho + \alpha$ il logaritmo del valore massimo di $|E^*|$, β e γ rappresentano due parametri di forma ed infine f_{red} rappresenta la frequenza ridotta. Le curve maestre realizzate per le miscele AC11, AC22, AC32 e SMA11 sono state graficamente rappresentate in Figura 9.3. Per ciascuna miscela e temperatura di riferimento, i valori dei fattori di spostamento $\log \alpha(T)$ necessari per determinare la corrispondente frequenza ridotta sono stati riportati in Tabella 9.2.

Ancora nel 1996, sulla base di un'ampia serie di osservazioni sperimentali (1430 dati riferiti a 149 conglomerati bituminosi convenzionali), Witzak e Fonseca hanno studiato le relazioni che sussistono tra i parametri della funzione sigmoidale e le proprietà fisico-volumetriche dei conglomerati bituminosi. Ne è stata dunque ricavata la seguente equazione empirica:

Tabella 9.2: Fattori di spostamento $\log \alpha(T)$.

Miscela	Temperatura (°C)						
	45	30	20	15	10	0	-15
AC11_70_100	- 9.8	- 5.1	- 1.8	-	1.8	5.7	12.1
AC11_PmB45_80_50	- 9.9	- 5.2	- 1.8	-	1.8	5.7	12.2
AC11_PmB45_80_65	- 9.9	- 5.2	- 1.8	-	1.8	5.7	12.2
AC22_mB160_220FT	- 9.5	- 5.0	- 1.7	0.0	-	5.5	11.7
AC22_PmB45_80_65	- 10.4	- 6.3	- 3.2	-	0.0	3.5	9.2
AC22_50_70	- 11.4	- 6.8	- 3.5	-	0.0	3.8	10.0
AC32_50_70	- 9.2	- 4.8	- 1.7	0.0	-	5.3	11.3
SMA11_70_100	- 9.9	- 5.2	- 1.8	-	1.8	5.7	12.2
SMA11_PmB45_80_65	- 9.4	- 4.9	- 1.7	-	1.8	5.5	11.6

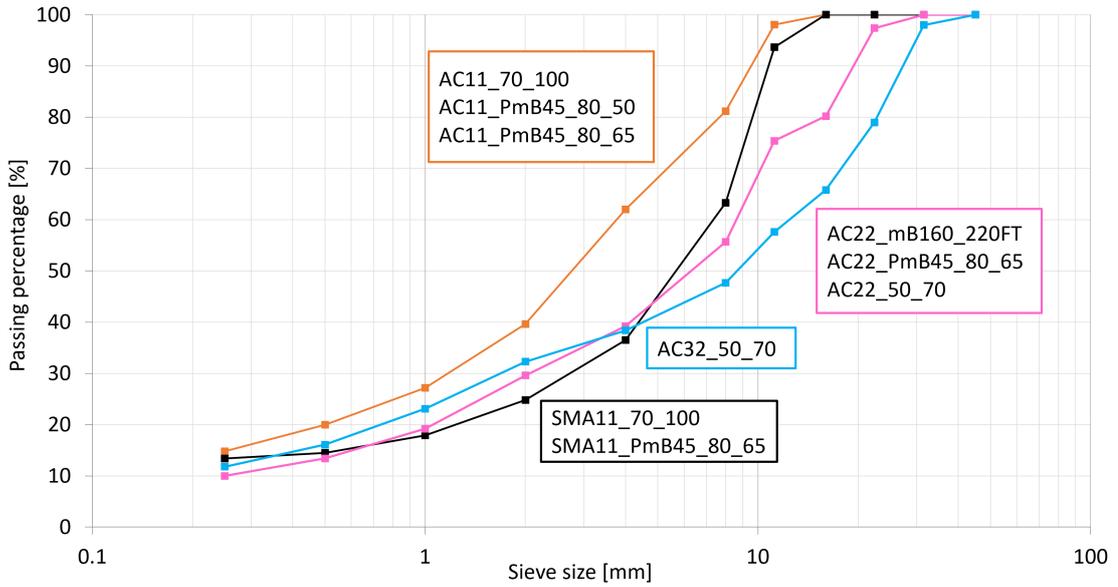


Figura 9.2: Curve granulometriche delle miscele analizzate.

$$\begin{aligned}
 \log|E^*| = & -0.261 + 0.008225 \cdot \rho_{No.200} - 0.00000101 \cdot (\rho_{No.200})^2 + 0.00196 \cdot \rho_{No.4} \\
 & - 0.03157 \cdot AV - 0.415 \cdot \frac{V_{b_{eff}}}{V_{b_{eff}} + AV} \quad (9.2.2) \\
 + & \frac{1.87 + 0.002808 \cdot \rho_{No.4} + 0.0000404 \cdot \rho_{3/8} - 0.0001786 \cdot (\rho_{3/8})^2 + 0.0164 \cdot \rho_{3/4}}{1 + e^{(-0.603313 - 0.313551 \cdot \log(f_l) - 0.393532 \cdot \log(\eta_b))}}
 \end{aligned}$$

$|E^*|$ rappresenta il modulo dinamico del conglomerato bituminoso (espresso in 10^5 psi) mentre $\rho_{No.200}$ è la percentuale di aggregato passante il vaglio No.200. $\rho_{No.4}$, $\rho_{3/8}$ e $\rho_{3/4}$ sono rispettivamente le percentuali di aggregato trattenute al setaccio No.4, al vaglio di maglia pari a 3/4 di pollice ed a quello di maglia pari a 3/8 di pollice. AV rappresenta la percentuale di vuoti d'aria nella miscela, $V_{b_{eff}}$ è il contenuto effettivo di legante bituminoso, f_l rappresenta la frequenza di carico (Hz) ed infine η_b rappresenta la viscosità del bitume (espressa in 10^6 poises) calcolata come segue:

$$\log_{10} \log_{10} \eta_b = A + VTS \cdot \log_{10}(T_R) \quad (9.2.3)$$

con T_R a rappresentare la temperatura espressa in gradi Rankine, mentre A e VTS rappresentano parametri di regressione determinati in funzione dei valori di PG (*Performance Grade*) [AASHTO, 1993].

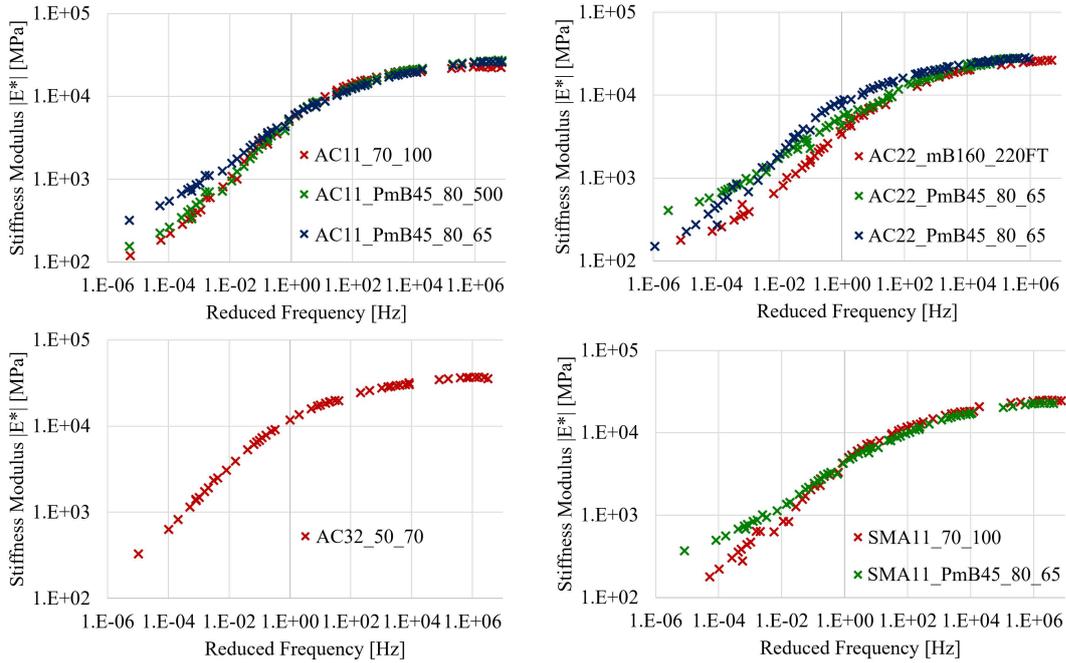


Figura 9.3: Curve maestre riferite alla temperatura di 15°C.

Nel 2006, Witczak et al. [Bari and Witczak, 2006] hanno combinato le osservazioni utilizzate nel modello originale di Fonseca e Witczak con ulteriori 1320 osservazioni aggiuntive relative a 56 ulteriori miscele (34 delle quali impiegavano leganti bituminosi modificati). Ne è stato derivato un nuovo modello denominato Witczak 1-37A, ancora oggi utilizzato negli Stati Uniti per prevedere i valori di modulo dinamico dei conglomerati bituminosi utilizzati nei progetti di livello 2 e 3 secondo il M-EPDG [AASHTO, 1993]. L'equazione predittiva è la seguente, dove i significati fisici delle variabili coinvolte sono analoghi a quelli precedentemente descritti:

$$\begin{aligned} \log|E^*| = & -1.249937 + 0.029232 \cdot \rho_{No.200} - 0.001767 \cdot (\rho_{No.200})^2 + 0.002841 \cdot \rho_{No.4} \\ & - 0.058097 \cdot AV - 0.802208 \cdot \frac{V_{b_{eff}}}{V_{b_{eff}} + AV} \\ & + \frac{3.871977 + 0.0021 \cdot \rho_{No.4} + 0.003958 \cdot \rho_{3/8} - 0.000017 \cdot (\rho_{3/8})^2 + 0.005470 \cdot \rho_{3/4}}{1 + e^{(-0.716 \cdot \log(f_i) - 0.7425 \cdot \log(\eta_b))}} \end{aligned} \quad (9.2.4)$$

9.3 Modellazione predittiva con approccio CatBoost

Per ottimizzare le prestazioni del modello CatBoost da implementare, è necessaria un'attenta identificazione dei suoi iperparametri, cioè di quei parametri i cui valori vengono impostati prima

dell'inizio del processo di apprendimento. In questa fase, assegnato l'algoritmo di addestramento, il processo di regolazione degli iperparametri mira ad ottimizzare la topologia del modello ed i parametri di addestramento. In primo luogo, è necessario dunque definire il numero di iterazioni, parametro rappresentativo del numero massimo di alberi decisionali che possono essere costruiti per risolvere il problema di apprendimento automatico considerato. Un ulteriore e significativo parametro da impostare corrisponde alla massima profondità degli alberi, definita come il numero massimo di suddivisioni. Tale valore deve essere scelto con massima cautela, poiché una profondità massima eccessivamente limitata comporta una modellazione rapida ma imprecisa, mentre una profondità massima eccessivamente elevata comporterà una modellazione molto più accurata ma probabilmente incline al fenomeno dell'*overfitting* [Witten and James, 2013]. Per limitare la possibilità che ciò accada, è stata implementata una tecnica di ricampionamento e validazione nota come *k-fold cross-validation* (già descritta nei paragrafi precedenti) ed il valore di *k* è stato posto pari a 5. Inoltre, è stato implementato un metodo di rilevamento dell'*overfitting* per interrompere l'addestramento qualora questo fenomeno venga individuato. Ancor prima di costruire ogni nuovo albero, l'algoritmo CatBoost verificherà quante iterazioni sono state eseguite dopo quella con il valore ottimale della funzione di perdita senza che si sia osservato un miglioramento significativo della funzione di perdita stessa. Se questo numero supera quello impostato come limite superiore per il rilevatore di *overfitting*, l'addestramento viene interrotto. Tale soglia è stata impostata al valore predefinito, ovvero 20. Infine, un ulteriore parametro da impostare con precisione è il tasso di apprendimento, utilizzato per ridurre il passo del gradiente durante l'addestramento. Individuati gli iperparametri da ottimizzare, è stata eseguita una esaustiva ricerca a griglia utilizzando Python 3.8.5. I domini di esplorazione ed i valori identificati come ottimali sono stati descritti in dettaglio nella Tabella 9.3. Dal momento che due variabili sono state simultaneamente predette, la funzione di perdita identificata come più adatta al fine di ottimizzare gli iperparametri è stata il MultiRMSE, la cui equazione risulta:

$$\text{MultiRMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \sum_{d=1}^D (y_{T,i,d} - y_{P,i,d})^2} \quad (9.3.1)$$

N e *D* rappresentano rispettivamente il numero complessivo di osservazioni ed il numero di variabili simultaneamente predette. Il modello più performante, ovvero quello che ha ottenuto il valore più basso di MultiRMSE, è caratterizzato da un numero massimo di iterazioni pari a 5000, una profondità massima pari a 6 ed un tasso di apprendimento pari a 0,05.

Tabella 9.3: Sintesi della ricerca a griglia.

Caratteristica	Griglia	Valore selezionato
Numero massimo di iterazioni	1000, 5000, 10000	5000
Massima profondità	4, 5, 6	6
Tasso di apprendimento	0.1, 0.05, 0.01	0.05
k-fold cross-validation	-	5
Rilevatore di overfitting	-	20
Funzione di perdita	-	MultiRMSE

9.3.1 Variabili di input ed output

L'approccio CatBoost è stato utilizzato per analizzare i risultati della campagna sperimentale e creare un modello in grado di prevedere simultaneamente sia il modulo dinamico che l'angolo di fase di 9 diversi conglomerati bituminosi. Per ciascuna miscela, 3 repliche di ciascun provino sono state testate a 6 diverse temperature di prova (7 nel caso di due delle nove miscele) e 10 frequenze di carico, generando così 1680 osservazioni complessive. Tuttavia, i dati mediati sulle 3 repliche sono stati effettivamente utilizzati ai fini modellistici, ottenendo così una banca dati composta da 560 osservazioni per ciascuna variabile. Le variabili di input implementate nel modello, in linea con le equazioni empiriche Witczak-Fonseca e Witczak 1-37A, fanno riferimento alle granulometrie, alle proprietà fisiche ed alle condizioni di prova. Queste sono state descritte nel dettaglio nelle prime 12 righe della Tabella 9.4: $\rho_{3/4}$, $\rho_{3/8}$, $\rho_{No.4}$, $\rho_{No.200}$, AV, $V_{b,eff}$, f_l , T, η_b , A e VTS. In input è stata aggiunta anche una variabile categorica, in grado di distinguere la natura dell'aggregato (calcare o diabase). Le variabili di output, invece, fanno riferimento al comportamento meccanico dei conglomerati e sono rappresentate dalle ultime due righe della Tabella 9.4, ovvero $|E^*|$ e φ .

Tabella 9.4: Descrizione statistica delle variabili considerate dal modello CatBoost.

Variabile	Descrizione	Unità di misura	Numero osserv.	Media	Dev. standard
Categorica	Natura dell'aggregato	-	560	-	-
$\rho_{3/4}$	Trattenuto al vaglio 3/4"	%	560	7.2	9.1
$\rho_{3/8}$	Trattenuto al vaglio 3/8"	%	560	25.9	12.5
$\rho_{No.4}$	Trattenuto al vaglio No.4	%	560	50.5	11.2
$\rho_{No.200}$	Passante al vaglio No.200	%	560	7.9	1.4
AV	Percentuale vuoti d'aria	%	560	3.7	0.4
$V_{b,eff}$	Contenuto effettivo di bitume	%	560	13.3	1.9
f_l	Frequenza di carico	Hz	560	13.0	12.6
T	Temperatura	°C	560	14.4	19.9
η_b	Viscosità bitume	10^6 Poise	560	213913.3	520927.3
A	Costante di regressione	-	560	10	0.5
VTS	Costante di regressione	-	560	-3.3	0.2
$ E^* $	Modulo dinamico	MPa	560	10859.5	9607.3
φ	Angolo di fase	°	560	21.8	13.8

9.3.2 Correlazioni esistenti tra le variabili

La matrice di correlazione di Pearson [Pallant, 2020] è stata utilizzata per esaminare in via preliminare l'intensità delle correlazioni esistenti tra le variabili considerate. Il coefficiente di correlazione di Pearson (R) tra due variabili è sempre compreso tra +1 e -1: il segno + è indice di una correlazione positiva (per cui se una variabile aumenta, si verifica anche un aumento dell'altra), mentre il segno - è indice di una correlazione negativa (per cui se una variabile aumenta, si verifica una diminuzione dell'altra). La dimensione del valore assoluto rappresenta

invece una misura dell'intensità della correlazione. Gli estremi dell'intervallo corrispondono ad una correlazione perfetta, il che significa che è possibile determinare il valore esatto di una variabile a partire dal valore dell'altra. Per contro, un valore di R pari a 0 significa che non esiste alcuna correlazione tra le due variabili: conoscere il valore di una variabile non aiuta a prevedere il valore dell'altra. I valori di R calcolati per ogni coppia di variabili sono riportati in una matrice di correlazione (Figura 9.4). Ai fini della modellazione predittiva, è utile notare che il modulo dinamico ha una forte correlazione negativa con la temperatura ($R=-0.90$) e una forte correlazione positiva con la viscosità del bitume ($R=+0.63$). Viceversa, l'angolo di fase ha una forte correlazione positiva con la temperatura ($R=+0.88$) e una forte correlazione negativa con la viscosità del bitume ($R=-0.53$).

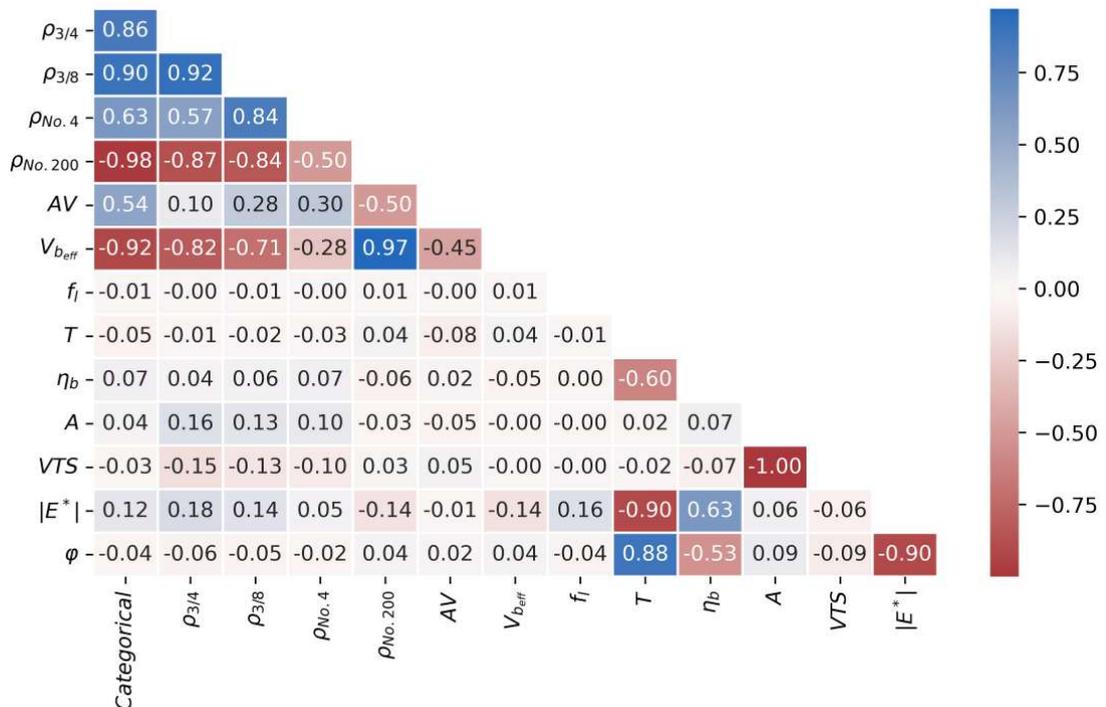


Figura 9.4: Matrice delle correlazioni di Pearson.

9.4 Discussione dei risultati

Prima di essere utilizzate come input del modello, tutte le variabili sono state normalizzate. Ogni osservazione di una specifica variabile è stata dunque mappata in un intervallo $[0, +1]$ i cui estremi inferiore e superiore corrispondono rispettivamente ai valori minimo e massimo assunti dalla variabile in oggetto. Questa procedura di pre-elaborazione dei dati è ben consolidata nel *Machine Learning* dal momento che contribuisce a ridurre i tempi computazionali e, al tempo stesso, ad aumentare sensibilmente l'efficienza dei modelli predittivi. Unitamente alla normaliz-

zazione, è stata implementata una *cross-validation* con k pari a 5 per ridurre la distorsione della valutazione delle performance predittive del modello. La banca dati complessiva è stata quindi suddivisa in due sottoinsiemi distinti in maniera casuale: un set di addestramento composto dall'80% delle osservazioni totali (448/560) ed uno di test composto dal 20% delle osservazioni totali (112/560). Tale scelta di partizionamento è risultata in linea con quanto riportato nella letteratura scientifica di riferimento [Hussain et al., 2021]. Il set di test è stato lasciato invariato, consentendo di testare il modello su una porzione di banca dati mai utilizzata, mentre il set di addestramento è stato riorganizzato come segue. È stato suddiviso in cinque partizionamenti differenti, in modo che quattro di questi potessero essere utilizzati per addestrare effettivamente il modello mentre quello rimanente sarebbe servito per validarlo. Il set di test è stato lasciato invariato, consentendo di testare il modello su una porzione di dati mai utilizzata, mentre il set di addestramento è stato riorganizzato come segue: è stato suddiviso in cinque parti differenti, in modo che quattro di esse potessero essere utilizzate per addestrare il modello e la rimanente per validarlo. Il processo di addestramento/validazione è stato ripetuto iterativamente 5 volte, in modo da poter variare ogni volta il partizionamento di validazione e monitorare un indice di performance (cioè un punteggio di validazione). La media dei cinque punteggi di validazione ha restituito le prestazioni effettive del modello per ciascuna iterazione di addestramento. L'andamento grafico della procedura di addestramento è stato riportato in Figura 9.5.

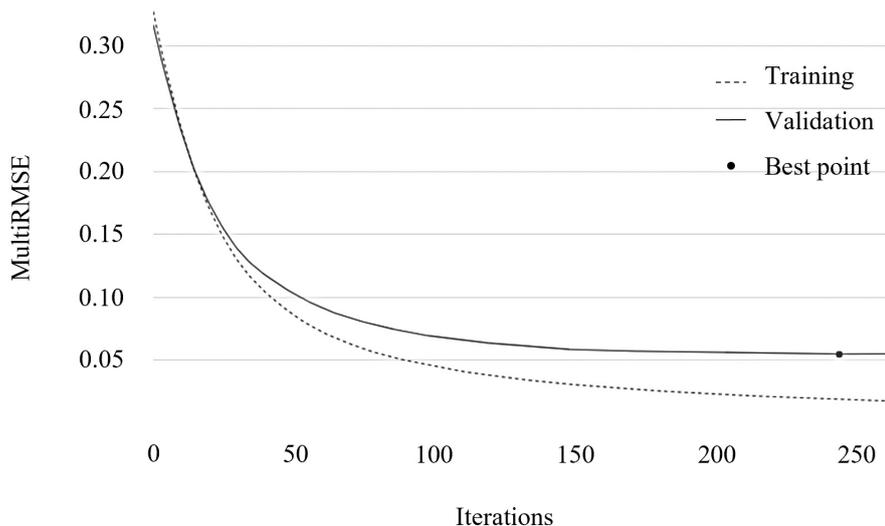


Figura 9.5: Rappresentazione del processo di addestramento del modello.

Si può osservare come all'inizio del processo l'errore sia piuttosto elevato e diminuisca progressivamente con il susseguirsi delle iterazioni. L'addestramento si è interrotto una volta raggiunta la convergenza, ossia quando non era più possibile osservare alcuna riduzione della funzione di perdita. Tuttavia, al diminuire della funzione di perdita di addestramento, è necessario porre molta attenzione all'andamento della funzione di perdita di validazione. Un contemporaneo aumento di quest'ultima, infatti, suggerirebbe il verificarsi del fenomeno dell'*overfitting*. Osservando la Figura 9.5, si può notare come sia la curva di addestramento

che quella di validazione decrescano con il procedere delle iterazioni. Inoltre, il rilevatore di *overfitting* impostato a 20 iterazioni ha permesso di individuare il miglior punto di validazione: oltre questa soglia non è più possibile apprezzare una diminuzione significativa della funzione di perdita di validazione ($|F^{(t+1)} - F^t| \leq 10^{-4}$). I risultati dell'addestramento possono essere riassunti come segue: rispetto al modulo dinamico $|E^*|$, MAE, RMSE e R^2 sono risultati rispettivamente pari a 279,68 MPa, 415,12 MPa e 0,998; per quanto riguarda l'angolo di fase φ , MAE, RMSE e R^2 sono risultati rispettivamente pari a 0,69°, 0,93° e 0,995. I parametri relativi alla migliore configurazione del modello sono stati dunque fissati e mantenuti costanti prima di procedere alla successiva fase di test.

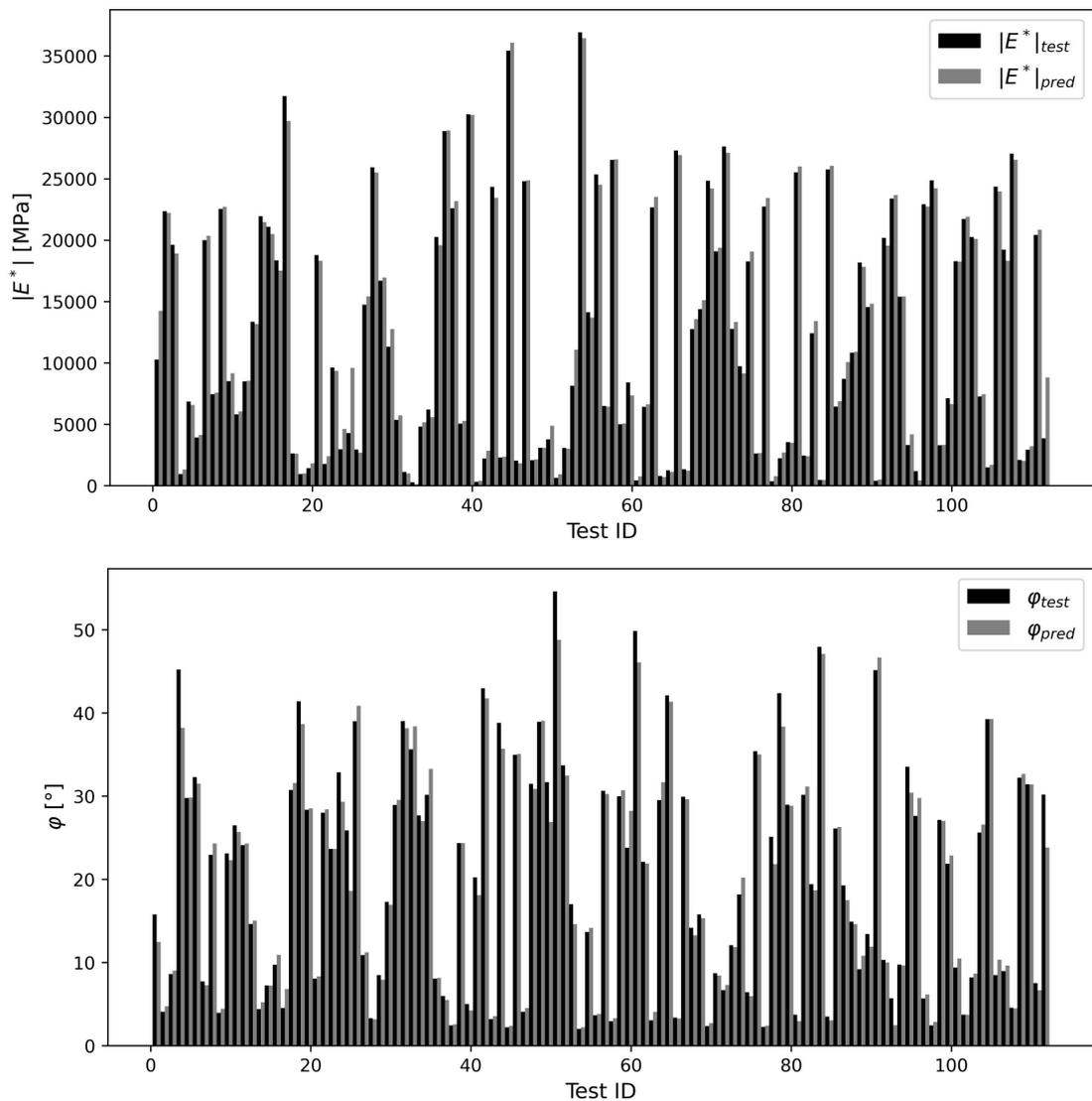
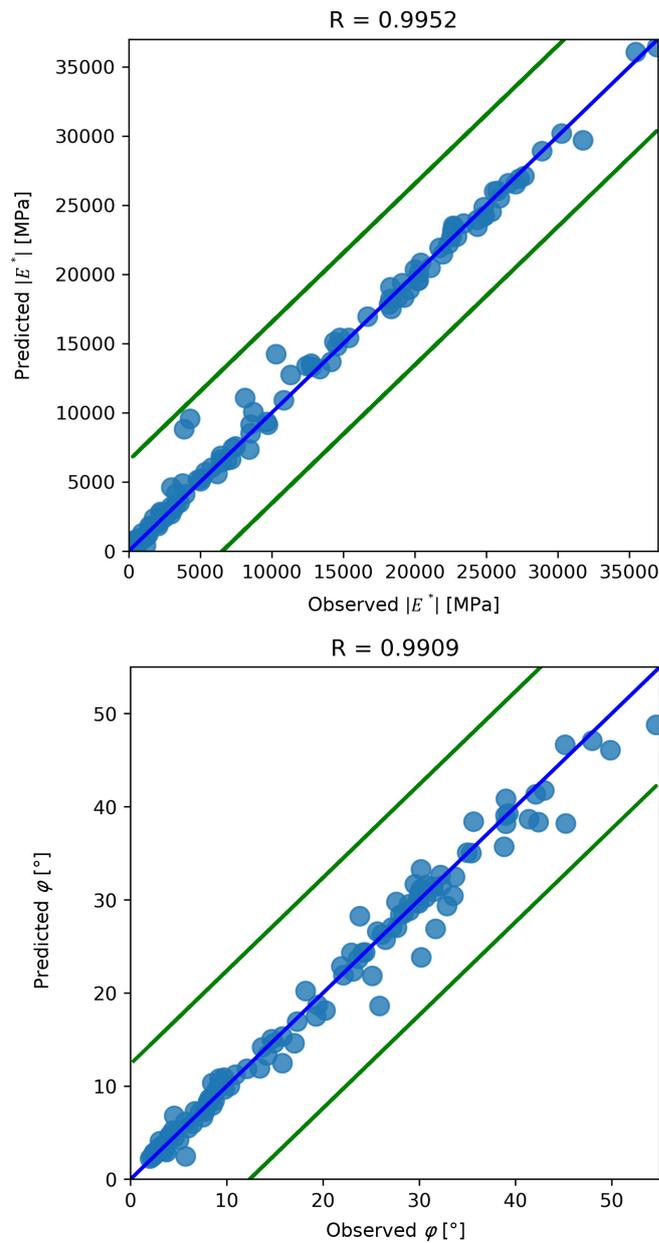


Figura 9.6: Valori di $|E^*|$ e φ sperimentalmente osservati e predetti dal modello CatBoost.

Figura 9.7: Grafici di regressione del modello CatBoost per $|E^*|$ e φ .

Al fine di ottenere una maggiore comprensione delle capacità predittive della metodologia sviluppata, la Figura 9.6 mostra i confronti tra i target sperimentali dell'insieme di test e gli output previsti dal modello CatBoost in riferimento ad entrambe le variabili, ovvero $|E^*|$ e φ . In questi grafici a istogramma, i valori delle osservazioni che compongono il set di test sono rappresentati in nero, mentre le rispettive previsioni effettuate dal modello CatBoost sono rappresentate in grigio. L'asse orizzontale è denominato Test ID e consente di identificare

ciascuna delle 112 coppie $|E^*|-\varphi$ che compongono l'insieme delle osservazioni di test. Si può apprezzare come le previsioni siano notevolmente prossime ai dati sperimentali osservati, per entrambi i parametri considerati. Questi risultati sono particolarmente significativi dal punto di vista ingegneristico poiché evidenziano l'affidabilità del modello CatBoost sviluppato e la bontà delle sue previsioni. Inoltre, dimostrano come tale modello sia in grado di prevedere simultaneamente e con elevata precisione due parametri fondamentali per la caratterizzazione del comportamento meccanico di diversi conglomerati bituminosi.

Per analizzare ancora più a fondo le performance predittive del modello, sono stati riportati due grafici di regressione, uno relativo al modulo dinamico $|E^*|$ e l'altro relativo all'angolo di fase φ (Figura 9.7). La linea continua blu inclinata a 45 gradi rappresenta la cosiddetta linea di uguaglianza, ossia la linea che riflette una previsione accurata al 100%, mentre i cerchi azzurri rappresentano le previsioni effettuate mediante il modello CatBoost. Si può osservare come le previsioni non si discostino molto dalla linea di uguaglianza, confermando i considerevoli risultati descritti dai coefficienti di correlazione di Pearson riportati in alto e riferiti alle coppie osservazione sperimentale-predizione del modello. Inoltre, tutte le previsioni di entrambe le variabili rientrano nei limiti di confidenza del 99%, rappresentati dalle linee continue verdi.

9.4.1 Analisi di sensitività

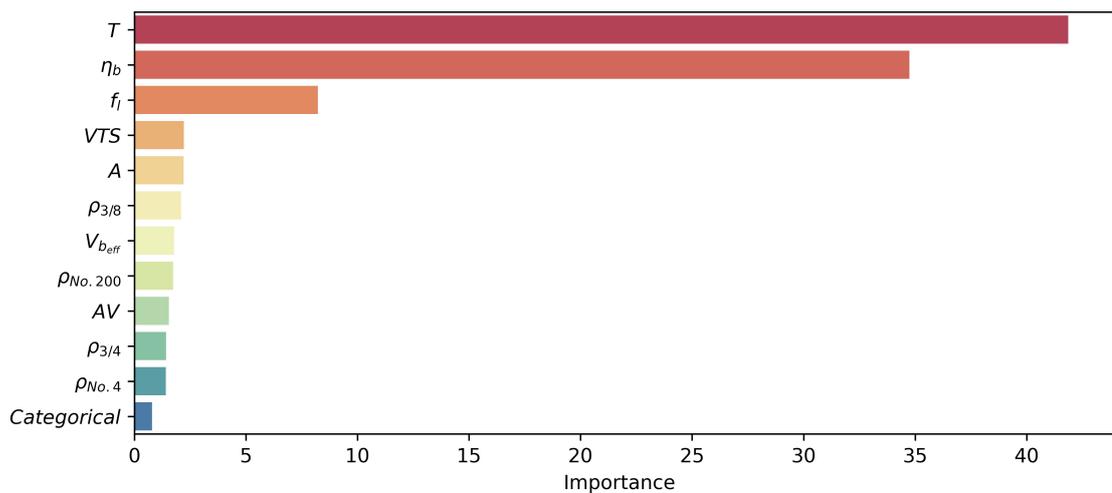


Figura 9.8: Stima dell'importanza di ciascuna variabile.

La stima dell'influenza di una determinata variabile sulle previsioni di un modello di apprendimento automatico è spesso complessa. Per questo motivo, utilizzando Python 3.8.5, è stato introdotto un algoritmo di misurazione dell'importanza delle variabili per determinare quanto cambiavano in media le predizioni del modello CatBoost al variare dei valori assunti dalle singole variabili. Quanto più elevato sarà il valore di importanza, tanto maggiore sarà la variazione del valore predetto in caso di modifica della rispettiva variabile. Inoltre, i valori di importanza sono stati normalizzati affinché la loro somma risultasse pari al 100%. Ciò è stato possibile poiché

i valori di importanza sono sempre non negativi. In Figura 9.8 vengono mostrati i risultati dell'analisi di sensitività effettuata, in cui l'importanza delle singole variabili viene mostrata come un elenco di coppie ID - importanza ordinate in base alla magnitudo dell'importanza. Si può notare come la temperatura (41,86%), la viscosità del bitume (34,74%) e la frequenza di carico (8,22%) risultino essere più critiche delle costanti empiriche di regressione, delle proprietà volumetriche e della natura degli aggregati (15,18% complessivamente) nella predizione dei parametri caratterizzanti il comportamento meccanico, ovverosia il modulo dinamico e l'angolo di fase.

9.4.2 Analisi SHAP

Per ottenere ulteriori informazioni sulle predizioni effettuate dal modello CatBoost, è stata eseguita anche un'analisi SHAP (*SHapley Additive exPlanations*). Tale procedura è stata sviluppata per comprendere i motivi per cui un modello produca determinati risultati, ponendo così l'accento sul concetto di interpretabilità delle predizioni, oltre che su quello più diffuso di accuratezza [Lundberg and Lee, 2017]. Siano f il modello CatBoost da interpretare e g il modello di interpretazione. Si assuma che $f(x)$ sia la predizione effettuata sulla base dell'input x . M rappresenti l'insieme di tutte le variabili, $z' \in \{0, 1\}^M$ sia pari ad 1 quando una variabile è in osservazione (in caso contrario sarà pari a 0), $|z'|$ sia il numero di elementi non nulli in z' , $z' \subseteq x$ rappresenti tutti i vettori z' in cui gli elementi non nulli sono un sottoinsieme degli elementi non nulli in x , e sia $z' \setminus i$ l'impostazione relativa a $z' = 0$ [Lundberg and Lee, 2017]. Il contributo fornito dalla variabile i -esima ϕ_i alla predizione dell'output del modello è determinato come segue:

$$\phi_i = \sum_{z' \subseteq x} \frac{|z'|!(M - |z'| - 1)!}{M!} [f_x(z') - f_x(z' \setminus i)] \quad (9.4.1)$$

g è definita come una funzione lineare di caratteristiche binarie [50] secondo il seguente metodo di attribuzione additiva delle caratteristiche:

$$g(z') = \phi_0 + \sum_{i=1}^M \phi_i z'_i \quad (9.4.2)$$

i valori ϕ_i sono noti come valori di Shapley [Shapley et al., 1953] e assegnano ad ogni variabile un'importanza specifica per ciascuna particolare previsione. In questo caso, sono state eseguite due analisi SHAP distinte: una per ciascun output del modello, ovverosia il modulo dinamico $|E^*|$ e l'angolo di fase φ . I risultati sono stati sintetizzati in figura 9.9 all'interno di diagrammi denominati *beeswarm* (letteralmente, sciame d'api). Per la predizione di ciascun output, le variabili sono state classificate dall'alto verso il basso in ordine decrescente di importanza. Si può notare ancora una volta come la temperatura rivesta la massima importanza sia nella predizione del modulo dinamico che in quella dell'angolo di fase, immediatamente seguita dalla viscosità del bitume e dalla frequenza di carico (a conferma di quanto già osservato in figura 9.8). Meno impattanti sono risultate le costanti empiriche di regressione. Infine, le proprietà volumetriche e la natura degli aggregati hanno dimostrato di rivestire un'importanza minore in fase predittiva. È possibile osservare che l'ordine di importanza in Figura 9.9 risulti

leggermente diverso da quello mostrato in Figura 9.8: ciò accade in virtù della separazione delle due analisi e del fatto che l'importanza in Figura 9.9 sia specificata per ciascun output. Tuttavia, oltre all'importanza delle variabili, i diagrammi SHAP aggiungono ulteriori informazioni (come suggerito dall'utilizzo dei colori): l'effetto delle variabili. A titolo esemplificativo, considerando la variabile temperatura ed il diagramma *beeswarm* riferito al modulo dinamico $|E^*|$, si può osservare come valori elevati di temperatura T (indicati in rosso) risultino in valori negativi di Shapley e, di conseguenza, in previsioni di valori più bassi di $|E^*|$. Tale tendenza si inverte gradualmente spostandosi lungo l'asse orizzontale dove valori di temperatura più bassi (indicati in blu) risultano invece in valori di Shapley positivi e, di conseguenza, in previsioni di valori più alti di $|E^*|$. L'identificazione di tali tendenze aiuta a comprendere la relazione esistente tra le singole variabili di input e lo specifico output. In generale, muovendosi da sinistra a destra, una transizione graduale dal blu al rosso denota una relazione di proporzionalità diretta tra target e variabile di input; al contrario, una transizione graduale dal rosso al blu denota una relazione di proporzionalità inversa tra le due grandezze. La prima è osservabile per le coppie $(|E^*|-\eta_b)$, $(|E^*|-f_i)$ e $(\varphi-T)$, mentre la seconda è osservabile per le coppie $(|E^*|-T)$, $(\varphi-\eta_b)$ e $(\varphi-f_i)$.

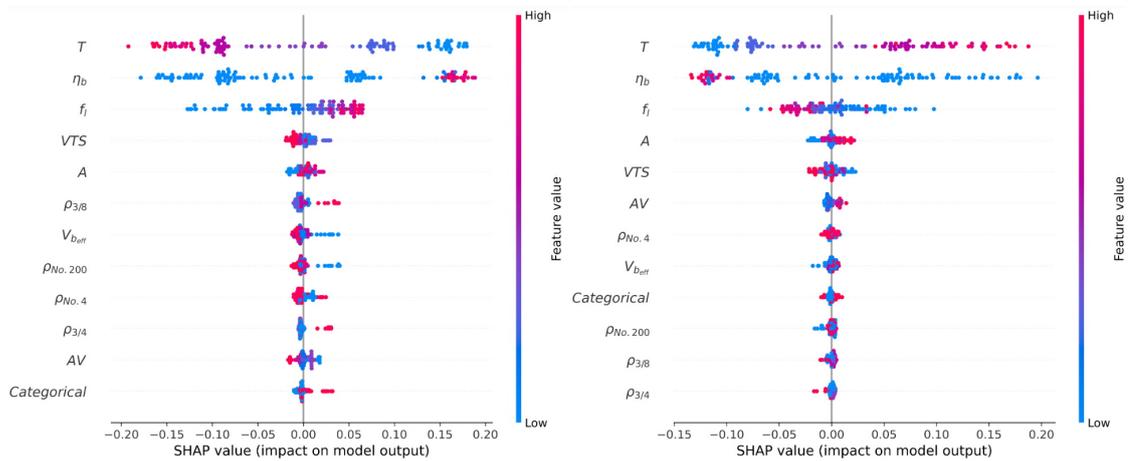


Figura 9.9: Diagrammi SHAP ottenuti per $|E^*|$ (sinistra) e φ (destra).

9.4.3 Confronto tra modelli

Un'approfondita analisi comparativa è stata eseguita per confrontare i risultati prodotti dai diversi modelli, ovvero i due basati sulle equazioni empiriche precedentemente descritte (Witczak-Fonseca e Witczak 1-37A), uno stato dell'arte delle reti neurali artificiali (*State-of-Art Artificial Neural Network* - SoA-ANN) ed il modello CatBoost basato sugli alberi decisionali. La descrizione particolareggiata della SoA-ANN implementata in MATLAB® è stata ricavata dal lavoro di [Baldo et al., 2022a]. Al fine di effettuare un confronto corretto ed equilibrato, entrambi i modelli di apprendimento automatico sono stati calibrati sulla base della stessa banca dati, utilizzando le stesse tecniche di pre-elaborazione e di ricampionamento. I quattro modelli sopracitati sono

stati quindi confrontati sulla base delle metriche di regressione descritte al paragrafo 3.8.2 ed i risultati sono stati sintetizzati in Tabella 9.5.

Tabella 9.5: Performance predittive dei diversi modelli analizzati.

Vettore di test $ E^* $	Witczak-Fonseca	Witczak 1-37A	SoA-ANN	CatBoost
MAE	3616.93	2146.30	640.05	552.21
MAPE	67.58	41.25	21.60	12.04
MSE	$2.03 \cdot 10^7$	$8.43 \cdot 10^6$	$6.88 \cdot 10^5$	$9.85 \cdot 10^5$
RMSE	4506.01	2904.13	829.21	992.36
R	0.9370	0.9694	0.9965	0.9952
R^2	0.7904	0.9129	0.9929	0.9898
Vettore di test φ	Witczak-Fonseca	Witczak 1-37A	SoA-ANN	CatBoost
MAE	-	-	1.05	1.19
MAPE	-	-	10.90	7.84
MSE	-	-	2.17	3.68
RMSE	-	-	1.47	1.92
R	-	-	0.9942	0.9908
R^2	-	-	0.9884	0.9804

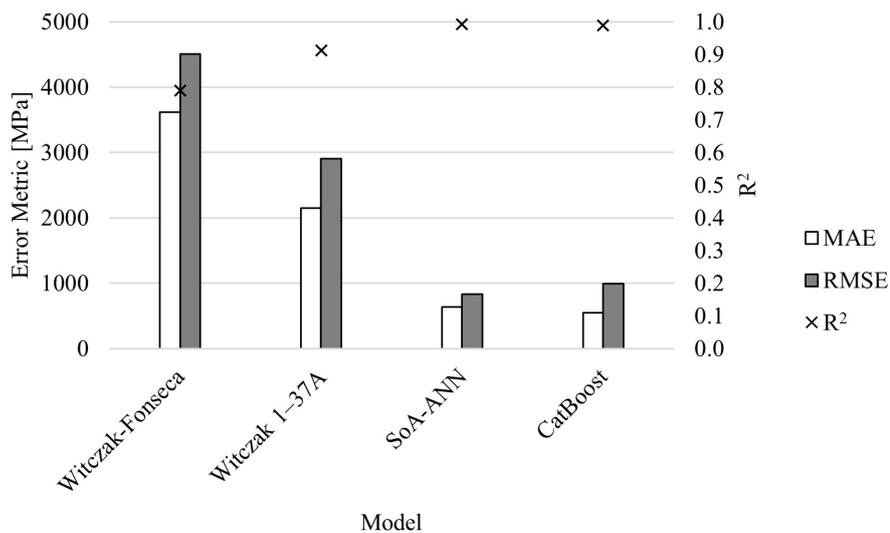


Figura 9.10: Rappresentazione grafica delle prestazioni predittive in termini di $|E^*|$.

Sulla base del vettore di test $|E^*|$ si possono dedurre le seguenti principali considerazioni. I modelli ML restituiscono prestazioni superiori a quelli basati sulla regressione, con CatBoost che ottiene sia il minimo MAE (552,21 MPa) che il minimo MAPE (12,04%). In termini delle altre quattro metriche di bontà di adattamento, il confronto tra i modelli SoA-ANN e CatBoost è molto competitivo e restituisce risultati confrontabili, con SoA-ANN che registra prestazioni

lievemente migliori (RMSE pari a 829,21 MPa). Confrontando solo i due modelli basati sulla regressione, Witczak 1-37A si comporta significativamente meglio di Witczak-Fonseca rispetto a tutte le sei metriche di performance implementate. Una rappresentazione sintetica dei risultati appena descritti può essere osservata in Figura 9.10. Tuttavia, è opportuno sottolineare un aspetto: sulla base delle stesse variabili di input, i modelli ML sono stati addestrati per prevedere simultaneamente non solo il modulo dinamico, ma anche l'angolo di fase. In questo modo, sono in grado di fornire una visione più dettagliata delle fondamentali proprietà viscoelastiche dei conglomerati bituminosi analizzati. Anche per quanto riguarda l'angolo di fase, le prestazioni di entrambi i modelli ML sono notevoli ed analogamente comparabili (Tabella 9.5) con MAE di circa 1° , MAPE inferiore all'11%, R e R^2 di circa 0,99 e 0,98, rispettivamente.

9.5 Conclusioni

Questo capitolo illustra una particolareggiata metodologia per l'implementazione di un innovativo algoritmo di *Machine Learning* basato sugli alberi decisionali, il CatBoost, per la previsione simultanea del modulo dinamico e dell'angolo di fase di nove diversi conglomerati bituminosi. Un'estesa campagna sperimentale è stata condotta utilizzando un apparato 4PBT e le osservazioni registrate sono state utilizzate per addestrare e successivamente testare il modello sviluppato. Sono state utilizzate 12 variabili di input, in riferimento alla natura ed alla granulometria degli aggregati, alla percentuale di vuoti d'aria, al contenuto effettivo ed alla viscosità del legante bituminoso, alla frequenza di carico ed alla temperatura di prova. Per valutare le performance del modello, sei diverse metriche di bontà di adattamento sono state implementate, ovvero MAE, MAPE, MSE, RMSE, R e R^2 . Un'approfondita analisi di sensitività è stata condotta per comprendere appieno il funzionamento del modello e l'influenza che ciascuna variabile ha esercitato sui due diversi parametri predetti. Le performance raggiunte dal modello sono state particolarmente soddisfacenti, risultato della notevole accuratezza delle previsioni effettuate. In termini di $|E^*|$, MAE, MAPE, MSE, RMSE, R e R^2 sono risultate rispettivamente pari a 552,21 MPa, 12,04%, $9,85 \times 10^5$ MPa², 992,36 MPa, 0,9952 e 0,9898. Analogamente, in termini di φ , le stesse metriche sono risultate rispettivamente pari a $1,19^\circ$, 7,84%, $3,68^\circ$, $1,92^\circ$, 0,9908 e 0,9804. Alla luce dei risultati appena descritti, è possibile trarre le seguenti conclusioni:

- Le variabili richieste come input dal modello predittivo possono essere facilmente determinate durante una procedura preliminare di progettazione delle miscele oppure facilmente derivate da esse, riducendo al minimo lo sforzo dei tecnici di laboratorio;
- Il modello sviluppato è in grado di fornire simultaneamente previsioni estremamente accurate sia del modulo dinamico che dell'angolo di fase, evitando la necessità che ulteriori test di laboratorio lunghi e onerosi debbano essere effettuati per poterli determinare sperimentalmente;
- In questo modo, parametri prestazionali come $|E^*|$ e φ possono essere facilmente stimati e quindi implementati nelle più comuni procedure di progettazione delle pavimentazioni in conglomerato bituminoso;

- Entrambi i metodi basati su algoritmi di *Machine Learning* hanno ottenuto risultati superiori a quelli prodotti dalle equazioni empiriche implementate, mostrando prestazioni di gran lunga migliori in termini di tutte le metriche di bontà di adattamento utilizzate;
- In particolare, i modelli CatBoost e SoA-ANN sono risultati competitivi e hanno evidenziato prestazioni pressoché equivalenti. Tuttavia, il modello CatBoost può essere facilmente implementato in Python e questo, unito alla sua elevata interpretabilità, lo rende preferibile al modello SoA-ANN.

La metodologia delineata è stata calibrata sui dati della campagna sperimentale descritta al paragrafo 9.2.1. Per gli sviluppi futuri, potrebbe essere facilmente adottata per analizzare anche dati più ampie e variegati, al fine di ottenere uno strumento sempre più potente e performante, in grado di fornire previsioni particolarmente accurate su una gamma di condizioni sempre più ampia. Infine, il modello CatBoost è stato sviluppato ed utilizzato per prevedere il modulo dinamico e l'angolo di fase dei conglomerati bituminosi, raggiungendo performance degne di nota. Tuttavia, la sua flessibilità potrebbe ispirare molteplici ulteriori applicazioni nell'ingegneria delle pavimentazioni, consentendo di studiare e modellare una vastità di fenomeni anche molto diversi, come la sensibilità all'acqua, le deformazioni permanenti e la resistenza alla fatica. È evidente che qualsiasi ulteriore applicazione del modello sviluppato diversa da quella proposta necessita di essere approfonditamente esplorata, introducendo nuove calibrazioni e ricercando i migliori iperparametri.

Capitolo 10

Conclusioni

” *In literature and in life we ultimately pursue, not conclusions, but beginnings.*

– Sam Tanenhaus

Il presente elaborato discute l'analisi e la modellazione del comportamento fisico-meccanico di miscele bituminose impiegate per la realizzazione di pavimentazioni stradali e/o aeroportuali mediante innovativi approcci di *machine learning*. Tale lavoro può essere contestualizzato all'interno del ben più ampio processo di transizione digitale legata al settore dell'ingegneria delle pavimentazioni poiché valuta la possibilità di impiegare moderni algoritmi di apprendimento automatico per analizzare banche dati relative a conglomerati bituminosi anche molto diversi tra loro, al fine di prevedere uno o (simultaneamente) più parametri legati ai fenomeni investigati. Nello specifico, tali modelli hanno permesso di migliorare significativamente efficacia e sostenibilità delle consuete procedure sperimentali di *mix-design* caratteristiche della tecnologia stradale e/o aeroportuale, in particolare riducendo drasticamente il carico di lavoro in laboratorio. Le principali conclusioni tratte da questo elaborato possono essere sintetizzate come segue:

- Gli algoritmi di apprendimento automatico possono essere implementati con successo anche nel settore dell'ingegneria delle pavimentazioni, permettendo la realizzazione di utili ed incoraggianti applicazioni in termini di sostenibilità delle procedure di progettazione dei conglomerati bituminosi;
- L'utilizzo di questi algoritmi, oltre a confermare quanto già noto sulle correlazioni esistenti tra variabili compositive e parametri prestazionali, permette di approfondire maggiormente anche eventuali relazioni meno evidenti che intercorrono tra tali variabili ed il comportamento fisico-meccanico dei corrispondenti conglomerati bituminosi;
- Le analisi di sensitività hanno permesso di individuare, in riferimento ai fenomeni investigati, le variabili maggiormente impattanti sulle prestazioni offerte dai conglomerati bituminosi in esame. Pertanto, dal momento che la maggioranza degli studi in questo

settore rappresenta il risultato di lunghe e costose campagne sperimentali, è raccomandabile raccogliere informazioni più dettagliate in termini delle variabili individuate come maggiormente significative, ottimizzando i benefici raggiungibili mediante modellazione predittiva;

- Ai fini di una stima accurata della/e variabile/i di output, è opportuno che sia disponibile un sufficiente quantitativo di osservazioni per l'intero intervallo di previsione. Infatti, se l'insieme di addestramento non dovesse contenere dati relativi ad un intervallo specifico, il modello, con molta probabilità, mostrerà maggiori difficoltà nell'effettuare una stima accurata dei valori dello specifico parametro in quel determinato intervallo. Ne consegue che l'utilizzo di una banca dati ampia e diversificata permetta di realizzare modelli predittivi dotati di superiori capacità di generalizzazione;
- Le tecniche di *data augmentation* possono giocare un ruolo cruciale nello sviluppo dei modelli predittivi, soprattutto se basati su banche dati di dimensioni limitate. Infatti, la generazione di dati sintetici che non alterino il significato intrinseco contenuto all'interno delle osservazioni sperimentali contribuisce ad incrementare sensibilmente la precisione delle previsioni effettuate dal modello di apprendimento automatico;
- Gli approcci basati sulla convalida incrociata restituiscono una valutazione delle prestazioni maggiormente attendibile rispetto a quanto accadrebbe con un semplice partizionamento casuale della banca dati. Quest'ultima procedura, infatti, genera una valutazione distorta delle prestazioni, significativamente influenzata da quali osservazioni abbiano costituito l'insieme di addestramento piuttosto che quello di test;
- L'ottimizzazione degli iperparametri rappresenta una fase cruciale dello sviluppo di un modello predittivo. Qualunque sia la tecnica di ottimizzazione implementata (ricerca casuale, a griglia, ottimizzazione Bayesiana, ecc.), è sempre opportuno accertarsi che l'insieme degli iperparametri utilizzati rappresenti la miglior configurazione possibile. In caso contrario, la qualità del modello potrebbe risultare drasticamente compromessa;
- Le prestazioni offerte dai modelli basati su approcci di *machine learning* sono risultate, nei contesti analizzati nell'ambito di questo elaborato, nettamente superiori rispetto a quelle restituite dall'utilizzo delle equazioni empiriche convenzionalmente adottate;
- Addestrate a partire dalle medesime banche dati, le metodologie basate sulle reti neurali artificiali e quelle basate sugli alberi decisionali con algoritmo CatBoost hanno restituito delle prestazioni molto soddisfacenti e pressoché confrontabili in termini di attendibilità delle previsioni. Tuttavia, la maggiore interpretabilità del modello CatBoost unita alla scrittura del suo codice sorgente tramite linguaggio di programmazione Python (ben più diffuso e facilmente implementabile) lo rende preferibile rispetto al contendente modello neurale sviluppato utilizzando un linguaggio di programmazione Matlab.
- Le metodologie sviluppate sono state calibrate sulla base delle descritte campagne sperimentali. Pertanto, qualsiasi applicazione dei modelli implementati diversa da quelle proposte in questo elaborato necessita di essere approfonditamente valutata, richiedendo

che vengano eseguite nuove procedure di addestramento, validazione e test oltre che di ottimizzazione degli iperparametri coinvolti.

Gli obiettivi di questo lavoro di ricerca non si limitavano alla sola produzione di interessanti risultati in relazione alla modellazione predittiva del comportamento fisico-meccanico di miscele bituminose per infrastrutture viarie ma si proponevano anche di sottolineare l'importanza dell'adozione di moderne tecniche "intelligenti" nel settore dell'ingegneria delle pavimentazioni. A prescindere dallo specifico campo di applicazione, infatti, i dati e le informazioni che ci circondano stanno aumentando in maniera rapida e prorompente, evidenziando la necessità di disporre di innovativi strumenti "intelligenti" in grado di gestirli ed utilizzarli in maniera efficiente e produttiva. La condivisione su larga scala delle conoscenze in materia di apprendimento automatico e la cooperazione internazionale tra esperti, ricercatori ed autorità competenti permetterebbero di raggiungere risultati ancor più positivi, dando il via ad una vera e propria rivoluzione digitale in ambito stradale e/o aeroportuale. La disponibilità di banche dati particolarmente ampie e diversificate permetterebbe di sviluppare strumenti dalle straordinarie potenzialità, frutto di una modellazione "intelligente" ponderata e di successo. Pertanto, l'auspicio per il prossimo futuro è che ci sia la possibilità di investire sempre più in progetti che favoriscano l'implementazione e l'integrazione di strumenti di apprendimento automatico all'interno del settore dell'ingegneria delle pavimentazioni.

Elenco delle figure

2.1	Successione di strati in una pavimentazione flessibile.	5
2.2	Composizione interna di un conglomerato bituminoso.	13
3.1	Rapporto tra AI, machine learning e deep learning.	25
3.2	Principali strategie di apprendimento automatico ed esempi applicativi.	27
3.3	Diagramma di flusso delle operazioni caratterizzanti un contesto ML.	30
3.4	Struttura generale di un modello predittivo di apprendimento automatico.	33
3.5	Diagramma a bersaglio.	36
3.6	Rappresentazione grafica del compromesso Bias-varianza.	37
3.7	Curve ideali di errore in fase di addestramento e validazione.	38
3.8	Tre degli algoritmi più diffusi per l'ottimizzazione degli iperparametri.	47
3.9	Schematizzazione dell'ottimizzazione Bayesiana degli iperparametri.	48
3.10	Schematizzazione grafica degli elementi valutati nella Precisione-recupero.	51
4.1	Principio di funzionamento di un SVR.	56
4.2	Schematizzazione di un neurone biologico.	58
4.3	Rappresentazione delle connessioni sinaptiche tra neuroni.	59
4.4	Stadi del potenziale d'azione.	60
4.5	Schematizzazione di un neurone artificiale.	61
4.6	Funzioni di attivazione parzialmente differenziabili ($\alpha = 1$).	63
4.7	Funzioni di attivazione totalmente differenziabili ($\beta = 1$).	64
4.8	Esempio di una rete feedforward a singolo strato.	66
4.9	Esempio di una rete feedforward multistrato.	66
4.10	Esempio di una rete ricorrente.	67
4.11	Schematizzazione dei passaggi dell'algoritmo di retropropagazione.	68
4.12	Notazione utile alla derivazione dell'algoritmo di retropropagazione.	69
4.13	Schematizzazione della struttura di un albero decisionale.	77
4.14	Esempio di albero decisionale binario con due variabili di input.	78
4.15	Esempio di albero decisionale obliquo tratto da [Lou and Obukhov, 2017].	87
5.1	Schematizzazione approccio di ricerca.	89
6.1	Schematizzazione di una prova tramite HWD.	91
6.2	Disposizione delle piste di volo dell'Aeroporto di Palermo.	93

ELENCO DELLE FIGURE

6.3	Punti di impatto individuati sulla pista di volo 02/20.	94
6.4	Mappatura delle deflessioni δ_0 registrate nei punti di impatto.	95
6.5	Spessore degli strati di pavimentazione ottenuti mediante carotaggi.	97
6.6	Mappatura dei moduli di rigidità dello strato superficiale calcolati a ritroso.	97
6.7	Funzioni di attivazione ELU, ReLU, TanH e LogS.	98
6.8	Punti di impatto sintetici ottenuti tramite Data Augmentation.	102
6.9	Metriche di performance del modello LogS-SNN.	104
6.10	Rappresentazione delle performance del modello LogS-SNN per ciascun fold.	105
6.11	Mappatura dei moduli di rigidità predetti dal modello LogS-SNN.	106
7.1	Curve granulometriche delle miscele BM12.5 preparate in laboratorio.	112
7.2	Curve granulometriche delle miscele BM20 preparate in laboratorio.	113
7.3	Curve granulometriche delle miscele BM20 preparate in impianto.	113
7.4	Strumentazione di laboratorio per test ITSM.	114
7.5	Metodologia descritta <i>step-by-step</i>	117
7.6	Andamento del coefficiente R_{CV} valutato durante le 500 iterazioni del BO.	118
7.7	Risultati della regressione riferiti ai 5 gruppi della convalida incrociata.	120
7.8	Confronto tra target sperimentali ed output previsti relativo al <i>fold</i> 4.	121
8.1	Curva granulometrica dell'aggregato progettata secondo requisiti MoRTH.	128
8.2	Esempi di RHA (a), BD (b), MD (c), SD (d), FA (e), OPC (f), LD (g) e SF (h).	129
8.3	Dati sperimentali (in nero) e dati sintetici (in rosso) per le miscele RHA.	133
8.4	AV in funzione di Fc per conglomerati con filler convenzionali e di riciclo.	135
8.5	VMA in funzione di Fc per conglomerati con filler convenzionali e di riciclo.	135
8.6	VFB in funzione di Fc per conglomerati con filler convenzionali e di riciclo.	136
8.7	MS in funzione di Fc per conglomerati con filler convenzionali e di riciclo.	137
8.8	MQ in funzione di Fc per conglomerati con filler convenzionali e di riciclo.	137
8.9	ITS in funzione di Fc per conglomerati con filler convenzionali e di riciclo.	138
8.10	ITSR in funzione di Fc per conglomerati con filler convenzionali e di riciclo.	138
8.11	CL in funzione di Fc per conglomerati con filler convenzionali e di riciclo.	139
8.12	Matrice delle correlazioni di Pearson.	141
8.13	Processo di addestramento del modello CatBoost.	142
8.14	Risultati dell'analisi di sensitività.	143
8.15	Confronti tra test e predizioni CatBoost per ciascuna variabile obiettivo.	145
8.16	Grafici di regressione per ciascuna variabile obiettivo.	146
9.1	Rappresentazione schematica della metodologia seguita.	150
9.2	Curve granulometriche delle miscele analizzate.	152
9.3	Curve maestre riferite alla temperatura di 15°C.	153
9.4	Matrice delle correlazioni di Pearson.	156
9.5	Rappresentazione del processo di addestramento del modello.	157
9.6	Valori di $ E^* $ e φ sperimentalmente osservati e predetti dal modello CatBoost.	158
9.7	Grafici di regressione del modello CatBoost per $ E^* $ e φ	159
9.8	Stima dell'importanza di ciascuna variabile.	160

ELENCO DELLE FIGURE

9.9	Diagrammi SHAP ottenuti per $ E^* $ (sinistra) e φ (destra).	162
9.10	Rappresentazione grafica delle prestazioni predittive in termini di $ E^* $	163

Elenco delle tabelle

2.1	Classi dimensionali ed apertura degli stacci a barre.	7
3.1	Alcune definizioni di Intelligenza Artificiale organizzate per categorie.	23
3.2	Vantaggi e svantaggi delle diverse strategie di apprendimento automatico.	29
3.3	Rappresentazione della procedura Leave-One-Out Cross-Validation.	42
3.4	Rappresentazione della procedura k-fold Cross-Validation.	43
3.5	Esempio di matrice di confusione.	49
6.1	Sintesi delle performance predittive dei migliori modelli neurali sviluppati.	104
7.1	Proprietà degli aggregati.	111
7.2	Proprietà dei bitumi.	112
7.3	Sintesi dei campioni Marshall preparati.	114
7.4	Caratteristiche meccaniche e volumetriche dei conglomerati investigati.	115
7.5	Valori di MSE e R ottenuti tramite ottimizzazione BO.	119
8.1	Proprietà dell'aggregato.	127
8.2	Proprietà del bitume.	130
8.3	Proprietà dei filler analizzati.	131
8.4	Riepilogo della ricerca a griglia.	132
8.5	Descrizione statistica delle variabili coinvolte nel modello predittivo.	140
8.6	Metriche di valutazione delle prestazioni del modello CatBoost.	141
9.1	Descrizione riassuntiva dei conglomerati bituminosi analizzati.	151
9.2	Fattori di spostamento $\log \alpha(T)$	151
9.3	Sintesi della ricerca a griglia.	154
9.4	Descrizione statistica delle variabili considerate dal modello CatBoost.	155
9.5	Performance predittive dei diversi modelli analizzati.	163

Bibliografia

- [AASHTO, 1993] AASHTO (1993). *AASHTO Guide for Design of Pavement Structures, 1993*, volume 1. AASHTO.
- [AASHTO T283, 2022] AASHTO T283 (2022). *Resistance of Compacted Asphalt Mixtures to Moisture-Induced Damage*. West Conshohocken, PA, USA.
- [Abdel-Wahed and Rashwan, 2016] Abdel-Wahed, T. A. and Rashwan, N. K. (2016). Application of cement dust and opc as mineral filler in the binder hot mix asphalt. In *LjMU International Conference on*.
- [Abdoli et al., 2015] Abdoli, M., Fathollahi, A., and Babaei, R. (2015). The application of recycled aggregates of construction debris in asphalt concrete mix design. *International Journal of Environmental Research*, 9(2).
- [Abreu et al., 2015] Abreu, L. P., Oliveira, J. R., Silva, H. M., and Fonseca, P. V. (2015). Recycled asphalt mixtures produced with high percentage of different waste materials. *Construction and Building Materials*, 84:230–238.
- [Adeli, 2001] Adeli, H. (2001). Neural networks in civil engineering: 1989–2000. *Computer-Aided Civil and Infrastructure Engineering*, 16(2):126–142.
- [Ahmed et al., 2017] Ahmed, T. M., Green, P. L., and Khalid, H. A. (2017). Predicting fatigue performance of hot mix asphalt using artificial neural networks. *Road Materials and Pavement Design*, 18(sup2):141–154.
- [Akbulut et al., 2012] Akbulut, H., Gürer, C., Çetin, S., and Elmacı, A. (2012). Investigation of using granite sludge as filler in bituminous hot mixtures. *Construction and Building Materials*, 36:430–436.
- [Akima, 1970] Akima, H. (1970). A new method of interpolation and smooth curve fitting based on local procedures. *Journal of the ACM (JACM)*, 17(4):589–602.
- [Akima, 1974] Akima, H. (1974). A method of bivariate interpolation and smooth surface fitting based on local procedures. *Communications of the ACM*, 17(1):18–20.
- [Aksoy et al., 2012] Aksoy, A., Iskender, E., and Kahraman, H. T. (2012). Application of the intuitive k-nn estimator for prediction of the marshall test (astm d1559) results for asphalt mixtures. *Construction and Building Materials*, 34:561–569.

- [Al-ammari et al., 2019] Al-ammari, M. A. S., Jakarni, F. M., Muniandy, R., and Hassim, S. (2019). The effect of aggregate and compaction method on the physical properties of hot mix asphalt. In *IOP Conference Series: Materials Science and Engineering*, volume 512, page 012003. IOP Publishing.
- [Alavi et al., 2011] Alavi, A. H., Ameri, M., Gandomi, A. H., and Mirzahosseini, M. R. (2011). Formulation of flow number of asphalt mixes using a hybrid computational method. *Construction and Building Materials*, 25(3):1338–1355.
- [Ali et al., 2021] Ali, Y., Hussain, F., Irfan, M., and Buller, A. S. (2021). An extreme gradient boosting model for predicting dynamic modulus of asphalt concrete mixtures. *Construction and Building Materials*, 295:123642.
- [Ali et al., 2017] Ali, Y., Irfan, M., Ahmed, S., and Ahmed, S. (2017). Empirical correlation of permanent deformation tests for evaluating the rutting response of conventional asphaltic concrete mixtures. *Journal of Materials in Civil Engineering*, 29(8):04017059.
- [Ali et al., 2015] Ali, Y., Irfan, M., Ahmed, S., Khanzada, S., and Mahmood, T. (2015). Sensitivity analysis of dynamic response and fatigue behaviour of various asphalt concrete mixtures. *Fatigue & Fracture of Engineering Materials & Structures*, 38(10):1181–1193.
- [Alrashydah and Abo-Qudais, 2018] Alrashydah, E. I. and Abo-Qudais, S. A. (2018). Modeling of creep compliance behavior in asphalt mixes using multiple regression and artificial neural networks. *Construction and Building Materials*, 159:635–641.
- [Althoey et al., 2023] Althoey, F., Akhter, M. N., Nagra, Z. S., Awan, H. H., Alanazi, F., Khan, M. A., Javed, M. F., Eldin, S. M., and Özkılıç, Y. O. (2023). Prediction models for marshall mix parameters using bio-inspired genetic programming and deep machine learning approaches: A comparative study. *Case Studies in Construction Materials*, 18:e01774.
- [Androjić and Marović, 2017] Androjić, I. and Marović, I. (2017). Development of artificial neural network and multiple linear regression models in the prediction process of the hot mix asphalt properties. *Canadian Journal of Civil Engineering*, 44(12):994–1004.
- [Arabani et al., 2017] Arabani, M., Tahami, S. A., and Taghipoor, M. (2017). Laboratory investigation of hot mix asphalt containing waste materials. *Road Materials and Pavement Design*, 18(3):713–729.
- [Ashtiani et al., 2021] Ashtiani, A., Murrell, S., and Brill, D. (2021). Machine learning approach to identifying key environmental factors for airfield asphalt pavement performance. In *Airfield and Highway Pavements 2021*, pages 328–337. American Society of Civil Engineers.
- [ASTM D4694-09, 2009] ASTM D4694-09 (2009). *Standard Test Method for Deflections with a Falling-Weight-Type Impulse Load Device*. West Conshohocken, PA, USA.
- [ASTM D4695-08, 2008] ASTM D4695-08 (2008). *Standard Guide for General Pavement Deflection Measurements*. West Conshohocken, PA, USA.

- [ASTM D6927-15, 2015] ASTM D6927-15 (2015). *Standard Test Method for Marshall Stability and Flow of Asphalt Mixtures*. West Conshohocken, PA, USA.
- [ASTM D6931-17, 2017] ASTM D6931-17 (2017). *Standard Test Method for Indirect Tensile (IDT) Strength of Asphalt Mixtures*. West Conshohocken, PA, USA.
- [Awed et al., 2022] Awed, A. M., Tarbay, E. W., El-Badawy, S. M., and Azam, A. M. (2022). Performance characteristics of asphalt mixtures with industrial waste/by-product materials as mineral fillers under static and cyclic loading. *Road Materials and Pavement Design*, 23(2):335–357.
- [Babagoli and Rezaei, 2022] Babagoli, R. and Rezaei, M. (2022). Prediction of moisture resistance of asphalt mastics modified by liquid anti stripping based on support vector regression, artificial neural network and kernel-based support vector regression methods. *Construction and Building Materials*, 335:127480.
- [Baduge et al., 2022] Baduge, S. K., Thilakarathna, S., Perera, J. S., Arashpour, M., Sharafi, P., Teodosio, B., Shringi, A., and Mendis, P. (2022). Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction*, 141:104440.
- [Baldo et al., 2018] Baldo, N., Manthos, E., and Pasetto, M. (2018). Analysis of the mechanical behaviour of asphalt concretes using artificial neural networks. *Advances in Civil Engineering*, 2018.
- [Baldo et al., 2021] Baldo, N., Miani, M., Rondinella, F., and Celauro, C. (2021). A machine learning approach to determine airport asphalt concrete layer moduli using heavy weight deflectometer data. *Sustainability*, 13(16):8831.
- [Baldo et al., 2022a] Baldo, N., Miani, M., Rondinella, F., Manthos, E., Valentin, J., et al. (2022a). Road pavement asphalt concretes for thin wearing layers: A machine learning approach towards stiffness modulus and volumetric properties prediction. *Periodica Polytechnica. Civil Engineering*, 66(4):1087–1097.
- [Baldo et al., 2022b] Baldo, N., Miani, M., Rondinella, F., Valentin, J., Vackcová, P., and Manthos, E. (2022b). Stiffness data of high-modulus asphalt concretes for road pavements: predictive modeling by machine-learning. *Coatings*, 12(1):54.
- [Bari, 2005] Bari, J. (2005). *Development of a new revised version of the Witczak E* predictive models for hot mix asphalt mixtures*. Arizona State University.
- [Bari and Witczak, 2006] Bari, J. and Witczak, M. (2006). Development of a new revised version of the witczak e* predictive model for hot mix asphalt mixtures (with discussion). *Journal of the Association of Asphalt Paving Technologists*, 75.
- [Barugahare et al., 2020] Barugahare, J., Amirkhanian, A. N., Xiao, F., and Amirkhanian, S. N. (2020). Predicting the dynamic modulus of hot mix asphalt mixtures using bagged trees ensemble. *Construction and Building Materials*, 260:120468.

- [Bashir et al., 2020] Bashir, D., Montañez, G. D., Sehra, S., Segura, P. S., and Lauw, J. (2020). An information-theoretic perspective on overfitting and underfitting. In *AI 2020: Advances in Artificial Intelligence: 33rd Australasian Joint Conference, AI 2020, Canberra, ACT, Australia, November 29–30, 2020, Proceedings 33*, pages 347–358. Springer.
- [Battiato et al., 1994] Battiato, G., Ame, E., and Wagner, T. (1994). Description and implementation of ro. ma. for urban road and highway network maintenance. In *3rd International Conference on Managing Pavement, San Antonio, TX*.
- [Battiti, 1992] Battiti, R. (1992). First-and second-order methods for learning: between steepest descent and newton’s method. *Neural computation*, 4(2):141–166.
- [Beale et al., 2010] Beale, M. H., Hagan, M. T., and Demuth, H. B. (2010). Neural network toolbox. *User’s Guide, MathWorks*, 2:77–81.
- [Behbahani et al., 2020] Behbahani, H., Hamed, G. H., and Gilani, V. N. M. (2020). Predictive model of modified asphalt mixtures with nano hydrated lime to increase resistance to moisture and fatigue damages by the use of deicing agents. *Construction and Building Materials*, 265:120353.
- [Behera and Mishra, 2012] Behera, B. and Mishra, M. K. (2012). Strength behaviour of surface coal mine overburden–fly ash mixes stabilised with quick lime. *International Journal of Mining, Reclamation and Environment*, 26(1):38–54.
- [Bellman, 1978] Bellman, R. (1978). *An Introduction to Artificial Intelligence: Can Computers Think?* Boyd & Fraser Publishing Company.
- [Bendel and Afifi, 1977] Bendel, R. B. and Afifi, A. A. (1977). Comparison of stopping rules in forward “stepwise” regression. *Journal of the American Statistical association*, 72(357):46–53.
- [Benesty et al., 2009] Benesty, J., Chen, J., Huang, Y., and Cohen, I. (2009). Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 1–4. Springer.
- [Bentéjac et al., 2021] Bentéjac, C., Csörgő, A., and Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54:1937–1967.
- [Bergstra et al., 2011] Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B. (2011). Algorithms for hyper-parameter optimization. In *25th annual conference on neural information processing systems (NIPS 2011)*, volume 24. Neural Information Processing Systems Foundation.
- [Bergstra and Bengio, 2012] Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2).
- [Bergstra et al., 2013] Bergstra, J., Yamins, D., and Cox, D. (2013). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *International conference on machine learning*, pages 115–123. PMLR.

- [Biligiri et al., 2010] Biligiri, K. P., Kaloush, K., and Uzan, J. (2010). Evaluation of asphalt mixtures' viscoelastic properties using phase angle relationships. *International Journal of Pavement Engineering*, 11(2):143–152.
- [Bohanec and Bratko, 1994] Bohanec, M. and Bratko, I. (1994). Trading accuracy for simplicity in decision trees. *Machine Learning*, 15:223–250.
- [Bouthillier et al., 2021] Bouthillier, X., Delaunay, P., Bronzi, M., Trofimov, A., Nichyporuk, B., Szeto, J., Mohammadi Sepahvand, N., Raff, E., Madan, K., Voleti, V., et al. (2021). Accounting for variance in machine learning benchmarks. *Proceedings of Machine Learning and Systems*, 3:747–769.
- [Breiman et al., 1984] Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). Classification and regression trees. *wadsworth int. Group*, 37(15):237–251.
- [Bretonnière, 1963] Bretonnière, S. (1963). Les déflectomètres à boulet pour l'étude des déflexions des chaussées sous charges dynamiques. *Bulletin des laboratoires des Ponts et Chaussées*.
- [Burmister, 1945] Burmister, D. M. (1945). The general theory of stresses and displacements in layered soil systems. iii. *Journal of applied Physics*, 16(5):296–302.
- [Bush III and Alexander, 1985] Bush III, A. J. and Alexander, D. R. (1985). Pavement evaluation using deflection basin measurements and layered theory. *Transport. Res.*, 1022:16–29.
- [Cai and Liu, 2020] Cai, Y. and Liu, X. (2020). Mechanical properties test of pavement base or subbase made of solid waste stabilized by acetylene sludge and fly ash. *AIP Advances*, 10(6).
- [Cao et al., 2016] Cao, P., Jin, F., Feng, D., Zhou, C., and Hu, W. (2016). Prediction on dynamic modulus of asphalt concrete with random aggregate modeling methods and virtual physics engine. *Construction and Building Materials*, 125:987–997.
- [Cao et al., 2020] Cao, P., Leng, Z., Shi, F., Zhou, C., Tan, Z., and Wang, Z. (2020). A novel viscoelastic damage model for asphalt concrete and its numerical implementation. *Construction and Building Materials*, 264:120261.
- [Carlucci et al., 2018] Carlucci, F., Cirà, A., and Coccorese, P. (2018). Measuring and explaining airport efficiency and sustainability: Evidence from italy. *Sustainability*, 10(2):400.
- [Ceylan et al., 2008] Ceylan, H., Gopalakrishnan, K., and Kim, S. (2008). Advanced approaches to hot-mix asphalt dynamic modulus prediction. *Canadian Journal of Civil Engineering*, 35(7):699–707.
- [Ceylan et al., 2009] Ceylan, H., Schwartz, C. W., Kim, S., and Gopalakrishnan, K. (2009). Accuracy of predictive models for dynamic modulus of hot-mix asphalt. *Journal of Materials in Civil Engineering*, 21(6):286–293.

- [Chang et al., 1992] Chang, D.-W., Kang, Y. V., Roesset, J. M., and Stokoe, K. (1992). Effect of depth to bedrock on deflection basins obtained with dynaflect and falling weight deflectometer tests. *Transportation Research Record*, pages 8–16.
- [Charniak and McDermott, 1985] Charniak, E. and McDermott, D. (1985). *Introduction to Artificial Intelligence*. Addison-Wesley series in computer science and information processing. Addison-Wesley.
- [Chehab et al., 2002] Chehab, G., Kim, Y., Schapery, R., Witzczak, M., and Bonaquist, R. (2002). Time-temperature superposition principle for asphalt concrete with growing damage in tension state. *Journal of the Association of Asphalt Paving Technologists*, 71.
- [Chen et al., 2003] Chen, D.-H., Bilyeu, J., Scullion, T., Lin, D.-F., and Zhou, F. (2003). Forensic evaluation of premature failures of texas specific pavement study-1 sections. *Journal of performance of constructed facilities*, 17(2):67–74.
- [Chen et al., 2011] Chen, M.-z., Lin, J.-t., Wu, S.-p., and Liu, C.-h. (2011). Utilization of recycled brick powder as alternative filler in asphalt mixture. *Construction and Building Materials*, 25(4):1532–1536.
- [Chen and Guestrin, 2016] Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.
- [Chern et al., 2005] Chern, S.-G., Lee, Y.-S., Hu, R.-F., and Chang, Y.-J. (2005). A research combines nondestructive testing and a neuro-fuzzy system for evaluating rigid pavement failure potential. *Journal of Marine Science and Technology*, 13(2):9.
- [Choudhary et al., 2022] Choudhary, J., Kumar, B., and Gupta, A. (2022). Performance evaluation of bauxite residue modified asphalt concrete mixes. *European Journal of Environmental and Civil Engineering*, 26(3):978–994.
- [Chrismer and Durham, 2010] Chrismer, J. and Durham, S. A. (2010). High volume fly ash concrete for highway pavements. In *Green Streets and Highways 2010: An Interactive Conference on the State of the Art and How to Achieve Sustainable Outcomes*, pages 390–400.
- [Christopher et al., 2010] Christopher, B. R., Schwartz, C. W., and Boudreau, R. L. (2010). *Geotechnical aspects of pavements: Reference manual*. US Department of Transportation, Federal Highway Administration.
- [Claessen et al., 1976] Claessen, A., Valkering, C., and Ditmarsch, R. (1976). Pavement evaluation with the falling weight deflectometer. In *Association of Asphalt Paving Technologists Proc*, volume 45.
- [Cominsky et al., 1994] Cominsky, R. J., Huber, G. A., Kennedy, T. W., and Anderson, M. (1994). *The superpave mix design manual for new construction and overlays*. SHRP-A-407. Strategic Highway Research Program Washington, DC, USA.

- [Cortes and Vapnik, 1995] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20:273–297.
- [Crovetti et al., 1989] Crovetti, J. A., Shahin, M., and Touma, B. E. (1989). Comparison of two falling weight deflectometer devices, dynatest 8000 and kuab 2m-fwd. In *Nondestructive Testing of Pavements and Backcalculation of Moduli*. ASTM International.
- [Da Silva et al., 2017] Da Silva, I. N., Spatti, D. H., Flauzino, R. A., Liboni, L. H. B., and dos Reis Alves, S. F. (2017). Artificial neural network architectures and training processes. In *Artificial neural networks*, pages 21–28. Springer.
- [Daneshvar and Behnood, 2022] Daneshvar, D. and Behnood, A. (2022). Estimation of the dynamic modulus of asphalt concretes using random forests algorithm. *International Journal of Pavement Engineering*, 23(2):250–260.
- [Darabi et al., 2019] Darabi, M. K., Huang, C.-W., Bazzaz, M., Masad, E. A., and Little, D. N. (2019). Characterization and validation of the nonlinear viscoelastic-viscoplastic with hardening-relaxation constitutive relationship for asphalt mixtures. *Construction and Building Materials*, 216:648–660.
- [Dash and Panda, 2018] Dash, S. S. and Panda, M. (2018). Influence of mix parameters on design of cold bituminous mix. *Construction and Building Materials*, 191:376–385.
- [de Oliveira Junior and de Farias, 2020] de Oliveira Junior, M. and de Farias, M. M. (2020). A simple numerical methodology to simulate creep and recovery tests in hma. *Construction and Building Materials*, 262:120793.
- [De Souza and De Almeida Filho, 2020] De Souza, N. M. and De Almeida Filho, A. T. (2020). A systematic airport runway maintenance and inspection policy based on a delay time modeling approach. *Automation in Construction*, 110:103039.
- [Demuth et al., 2014] Demuth, H. B., Beale, M. H., De Jess, O., and Hagan, M. T. (2014). *Neural network design*. Martin Hagan.
- [Di Benedetto et al., 2004] Di Benedetto, H., Olard, F., Sauzéat, C., and Delaporte, B. (2004). Linear viscoelastic behaviour of bituminous materials: From binders to mixes. *Road Materials and Pavement Design*, 5(sup1):163–202.
- [Di Benedetto et al., 2016] Di Benedetto, H., Sauzéat, C., and Clec'h, P. (2016). Anisotropy of bituminous mixture in the linear viscoelastic domain. *Mechanics of Time-Dependent Materials*, 20(3):281–297.
- [Dias et al., 2014] Dias, J. F., Picado-Santos, L., and Capitão, S. (2014). Mechanical performance of dry process fine crumb rubber asphalt mixtures placed on the portuguese road network. *Construction and Building Materials*, 73:247–254.
- [Erkens et al., 2002] Erkens, S., Liu, X., and Scarpas, A. (2002). 3d finite element model for asphalt concrete response simulation. *The International Journal Geomechanics*, 2(3):305–330.

- [Fahlman et al., 1988] Fahlman, S. E. et al. (1988). *An empirical study of learning speed in back-propagation networks*. Carnegie Mellon University, Computer Science Department Pittsburgh, PA, USA.
- [Fonseca and Witczak, 1996] Fonseca, O. and Witczak, M. W. (1996). A prediction methodology for the dynamic modulus of in-place aged asphalt mixtures (with discussion). *Journal of the Association of Asphalt Paving Technologists*, 65.
- [Foresee and Hagan, 1997] Foresee, F. D. and Hagan, M. T. (1997). Gauss-newton approximation to bayesian learning. In *Proceedings of international conference on neural networks (ICNN'97)*, volume 3, pages 1930–1935. IEEE.
- [Friedman, 2001] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.
- [Friedman et al., 2010] Friedman, J. H., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.
- [Gandomi et al., 2011] Gandomi, A. H., Alavi, A. H., Mirzahosseini, M. R., and Nejad, F. M. (2011). Nonlinear genetic-based models for prediction of flow number of asphalt mixtures. *Journal of Materials in Civil Engineering*, 23(3):248–263.
- [Garcia et al., 2014] Garcia, A., Norambuena-Contreras, J., Bueno, M., and Partl, M. N. (2014). Influence of steel wool fibers on the mechanical, thermal, and healing properties of dense asphalt concrete. *Journal of Testing and Evaluation*, 42(5):1107–1118.
- [García et al., 2015] García, S., Luengo, J., and Herrera, F. (2015). *Data preprocessing in data mining*. Springer.
- [Garg et al., 2004] Garg, N., Guo, E., and McQueen, R. (2004). Operational life of airport pavements. Technical report, U.S. Department of Transportation, Federal Aviation Administration.
- [Gelfand et al., 1989] Gelfand, S. B., Ravishankar, C., and Delp, E. J. (1989). An iterative growing and pruning algorithm for classification tree design. In *Conference Proceedings., IEEE International Conference on Systems, Man and Cybernetics*, pages 818–823. IEEE.
- [Ghafari et al., 2022] Ghafari, S., Ehsani, M., and Nejad, F. M. (2022). Prediction of low-temperature fracture resistance curves of unmodified and crumb rubber modified hot mix asphalt mixtures using a machine learning approach. *Construction and Building Materials*, 314:125332.
- [Gilani et al., 2020] Gilani, V. N. M., Hosseinian, S. M., Behbahani, H., and Hamed, G. H. (2020). Prediction and pareto-based multi-objective optimization of moisture and fatigue damages of asphalt mixtures modified with nano hydrated lime. *Construction and Building Materials*, 261:120509.

- [Giunta and Angela Pisano, 2006] Giunta, M. and Angela Pisano, A. (2006). One-dimensional visco-elastoplastic constitutive model for asphalt concrete. *Multidiscipline Modeling in Materials and Structures*, 2(2):247–264.
- [Goel and Das, 2008] Goel, A. and Das, A. (2008). Nondestructive testing of asphalt pavements for structural condition evaluation: a state of the art. *Nondestructive Testing and Evaluation*, 23(2):121–140.
- [Goktepe et al., 2006] Goktepe, A. B., Agar, E., and Lav, A. H. (2006). Advances in backcalculating the mechanical properties of flexible pavements. *Advances in engineering software*, 37(7):421–431.
- [Gopalakrishnan, 2012] Gopalakrishnan, K. (2012). Instantaneous pavement condition evaluation using non-destructive neuro-evolutionary approach. *Structure and Infrastructure Engineering*, 8(9):857–872.
- [Gopalakrishnan et al., 2009] Gopalakrishnan, K., Ceylan, H., and Guclu, A. (2009). Airfield pavement deterioration assessment using stress-dependent neural network models. *Structure and Infrastructure Engineering*, 5(6):487–496.
- [Gopalakrishnan and Thompson, 2005] Gopalakrishnan, K. and Thompson, M. R. (2005). Use of deflection basin parameters to characterize structural degradation of airport flexible pavements. In *Advances in Pavement Engineering*, pages 1–15. ASCE Library.
- [Gopalakrishnan and Thompson, 2007] Gopalakrishnan, K. and Thompson, M. R. (2007). Use of nondestructive test deflection data for predicting airport pavement performance. *Journal of transportation engineering*, 133(6):389–395.
- [Gordon and Tibshirani, 2012] Gordon, G. and Tibshirani, R. (2012). Karush-kuhn-tucker conditions. *Optimization*, 10(725/36):725.
- [Guo et al., 2016] Guo, Y., He, X., Peeta, S., and Weiss, W. J. (2016). Internal curing for concrete bridge decks: Integration of a social cost analysis in evaluation of long-term benefit. *Transportation Research Record*, 2577(1):25–34.
- [Haas et al., 1994] Haas, R., Hudson, W. R., and Zaniewski, J. P. (1994). *Modern pavement management*. Krieger Publishing.
- [Hagan and Menhaj, 1994] Hagan, M. T. and Menhaj, M. B. (1994). Training feedforward networks with the marquardt algorithm. *IEEE transactions on Neural Networks*, 5(6):989–993.
- [Hamed et al., 2021] Hamed, G. H., Esmaeeli, M. R., Najafi Moghaddam Gilani, V., and Hosseini, S. M. (2021). The effect of aggregate-forming minerals on thermodynamic parameters using surface free energy concept and its relationship with the moisture susceptibility of asphalt mixtures. *Advances in Civil Engineering*, 2021:1–15.
- [Han et al., 2021] Han, D., Yuan, M., and Hu, H. (2021). Establishing prediction master curve of dynamic modulus of asphalt mixture considering randomness of aggregate morphology. *Construction and Building Materials*, 294:123575.

- [Harnad, 1991] Harnad, S. (1991). Other bodies, other minds: A machine incarnation of an old philosophical problem. *Minds and Machines*, 1(1):43–54.
- [Harrington, 2012] Harrington, P. (2012). *Machine learning in action*. Simon and Schuster.
- [Hastie et al., 2009] Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.
- [Haugeland, 1985] Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. Cambridge: MIT Press.
- [Hebb, 1949] Hebb, D. O. (1949). The first stage of perception: growth of the assembly. *The Organization of Behavior*, 4(60):78–60.
- [Hellström et al., 2020] Hellström, T., Dignum, V., and Bensch, S. (2020). Bias in machine learning—what is it good for? *arXiv preprint arXiv:2004.00686*.
- [Hill, 1988] Hill, H. J. (1988). *Early life study of the FA409 full-depth asphalt concrete pavement sections*. PhD thesis, University of Illinois at Urbana-Champaign.
- [Hodgkin and Huxley, 1952] Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500.
- [Hoffman and Thompson, 1982] Hoffman, M. S. and Thompson, M. R. (1982). *Comparative study of selected nondestructive testing devices*, volume 852. Transportation Research Record.
- [Hossain and Zaniewski, 1991] Hossain, A. and Zaniewski, J. P. (1991). Characterization of falling weight deflectometer deflection basin. *Transportation Research Record*, 1293.
- [Hou et al., 2021] Hou, Y., Li, Q., Zhang, C., Lu, G., Ye, Z., Chen, Y., Wang, L., and Cao, D. (2021). The state-of-the-art review on applications of intrusive sensing, image processing techniques, and machine learning methods in pavement monitoring and analysis. *Engineering*, 7(6):845–856.
- [Huang et al., 2007] Huang, B., Shu, X., and Chen, X. (2007). Effects of mineral fillers on hot-mix asphalt laboratory-measured properties. *International Journal of Pavement Engineering*, 8(1):1–9.
- [Huang et al., 2022] Huang, Y., Shafiee, M., Charnley, F., and Encinas-Oropesa, A. (2022). Designing a framework for materials flow by integrating circular economy principles with end-of-life management strategies. *Sustainability*, 14(7):4244.
- [Huang et al., 2004] Huang, Y. H. et al. (2004). *Pavement analysis and design*, volume 2. Pearson Prentice Hall Upper Saddle River, NJ.
- [Hussain et al., 2021] Hussain, F., Ali, Y., Irfan, M., Ashraf, M., and Ahmed, S. (2021). A data-driven model for phase angle behaviour of asphalt concrete mixtures based on convolutional neural network. *Construction and Building Materials*, 269:121235.

- [ICAO, 2016] ICAO (2016). Annex 14: Aerodromes, volume i–aerodrome design and operations, 7th edition. *International Civil Aviation Organisation*.
- [James et al., 2013] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *Resampling Methods*, volume 112, chapter 5, pages 176–186. Springer.
- [Jang et al., 1997] Jang, J.-S. R., Sun, C.-T., and Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing - A Computational Approach to Learning and Machine Intelligence*. Prentice-Hall.
- [Justo-Silva et al., 2021] Justo-Silva, R., Ferreira, A., and Flintsch, G. (2021). Review on machine learning techniques for developing pavement performance prediction models. *Sustainability*, 13(9):5248.
- [Karbassi et al., 2014] Karbassi, A., Mohebi, B., Rezaee, S., and Lestuzzi, P. (2014). Damage prediction for regular reinforced concrete buildings using the decision tree algorithm. *Computers & Structures*, 130:46–56.
- [Ke et al., 2017] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- [Khajuria et al., 2022] Khajuria, A., Atienza, V. A., Chavanich, S., Henning, W., Islam, I., Kral, U., Liu, M., Liu, X., Murthy, I. K., Oyedotun, T. D. T., et al. (2022). Accelerating circular economy solutions to achieve the 2030 agenda for sustainable development goals. *Circular Economy*, 1(1):100001.
- [Khuntia et al., 2014] Khuntia, S., Das, A. K., Mohanty, M., and Panda, M. (2014). Prediction of marshall parameters of modified bituminous mixtures using artificial intelligence techniques. *International Journal of Transportation Science and Technology*, 3(3):211–227.
- [Kim and Kim, 2006] Kim, S.-H. and Kim, N. (2006). Development of performance prediction models in flexible pavement using regression analysis method. *KSCE Journal of Civil Engineering*, 10(2):91–96.
- [Komba et al., 2012] Komba, J. J., Maina, J., Anochie-Boateng, J., and O’Connell, J. S. (2012). Analytical modelling of visco-elastic behaviour of hot-mix asphalt. In *31st Southern African Transport Conference (SATC)*.
- [Kuhn et al., 2013] Kuhn, M., Johnson, K., et al. (2013). *Applied predictive modeling*, volume 26, chapter 4, pages 61–92. Springer.
- [Kuity et al., 2014] Kuity, A., Jayaprakasan, S., and Das, A. (2014). Laboratory investigation on volume proportioning scheme of mineral fillers in asphalt mixture. *Construction and Building Materials*, 68:637–643.
- [Kurzweil, 1990] Kurzweil, R. (1990). *The Age of Intelligent Machines*. MIT Press.

- [Kushner, 1964] Kushner, H. J. (1964). A New Method of Locating the Maximum Point of an Arbitrary Multippeak Curve in the Presence of Noise. *Journal of Basic Engineering*, 86(1):97–106.
- [Kutuk-Sert, 2016] Kutuk-Sert, T. (2016). Stability analyses of submicron-boron mineral prepared by mechanical milling process in concrete roads. *Construction and Building Materials*, 121:255–264.
- [Laurinavičius and Oginskas, 2006] Laurinavičius, A. and Oginskas, R. (2006). Experimental research on the development of rutting in asphalt concrete pavements reinforced with geosynthetic materials. *Journal of civil engineering and management*, 12(4):311–317.
- [Lee et al., 2016] Lee, B. X., Kjaerulf, F., Turner, S., Cohen, L., Donnelly, P. D., Muggah, R., Davis, R., Realini, A., Kieselbach, B., MacGregor, L. S., et al. (2016). Transforming our world: implementing the 2030 agenda through sustainable development goal indicators. *Journal of public health policy*, 37:13–31.
- [Lesueur et al., 2013] Lesueur, D., Petit, J., and Ritter, H.-J. (2013). The mechanisms of hydrated lime modification of asphalt mixtures: A state-of-the-art review. *Road materials and pavement design*, 14(1):1–16.
- [Li et al., 2012] Li, J., Zofka, A., and Yut, I. (2012). Evaluation of dynamic modulus of typical asphalt mixtures in northeast us region. *Road materials and pavement design*, 13(2):249–265.
- [Li and Wang, 2019] Li, M. and Wang, H. (2019). Development of ann-ga program for backcalculation of pavement moduli under fwd testing with viscoelastic and nonlinear parameters. *International Journal of Pavement Engineering*, 20(4):490–498.
- [Li et al., 2021] Li, Z., Guo, T., Chen, Y., Zhao, X., Chen, Y., Yang, X., and Wang, J. (2021). Road performance analysis of cement stabilized coal gangue mixture. *Materials Research Express*, 8(12):125502.
- [Liu et al., 2017] Liu, J., Yan, K., You, L., Liu, P., and Yan, K. (2017). Prediction models of mixtures' dynamic modulus using gene expression programming. *International Journal of Pavement Engineering*, 18(11):971–980.
- [Liu et al., 2019] Liu, L., Jiang, H., He, P., Chen, W., Liu, X., Gao, J., and Han, J. (2019). On the variance of the adaptive learning rate and beyond. *arXiv preprint arXiv:1908.03265*.
- [Liu and Wu, 2014] Liu, Q. T. and Wu, S. P. (2014). Effects of steel wool distribution on properties of porous asphalt concrete. In *Key Engineering Materials*, volume 599, pages 150–154. Trans Tech Publ.
- [Lou and Obukhov, 2017] Lou, Y. and Obukhov, M. (2017). Bdt: Gradient boosted decision tables for high accuracy and scoring efficiency. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1893–1901.

- [Lukjan et al., 2022] Lukjan, A., Iyaruk, A., and Somboon, C. (2022). Evaluation on mechanical deterioration of the asphalt mixtures containing waste materials when exposed to corrosion solutions. *International Journal of Engineering & Technology Innovation*, 12(2).
- [Lunardi, 2018] Lunardi, A. (2018). *Interpolation theory*, volume 16. Springer.
- [Lundberg and Lee, 2017] Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- [Lytton, 1989] Lytton, R. L. (1989). Backcalculation of layer moduli, state of the art. *Nondestructive testing of pavements and backcalculation of moduli*, pages 7–38.
- [MacKay, 1992] MacKay, D. J. (1992). Bayesian interpolation. *Neural computation*, 4:415–447.
- [Maimon and Rokach, 2005] Maimon, O. and Rokach, L. (2005). *Data mining and knowledge discovery handbook*, volume 2. Springer.
- [Majidifard et al., 2019] Majidifard, H., Jahangiri, B., Buttlar, W. G., and Alavi, A. H. (2019). New machine learning-based prediction models for fracture energy of asphalt mixtures. *Measurement*, 135:438–451.
- [Makovšek et al., 2012] Makovšek, D., Tominc, P., and Logožar, K. (2012). A cost performance analysis of transport infrastructure construction in slovenia. *Transportation*, 39:197–214.
- [Marcelino et al., 2021] Marcelino, P., de Lurdes Antunes, M., Fortunato, E., and Gomes, M. C. (2021). Machine learning approach for pavement performance prediction. *International Journal of Pavement Engineering*, 22(3):341–354.
- [Materials and Test Division Texas Department of Transportation, 2019] Materials and Test Division Texas Department of Transportation (2019). *Test Procedure for Cantabro Loss*. Austin, TX, USA.
- [Mazzoni et al., 2019] Mazzoni, G., Virgili, A., and Canestrari, F. (2019). Influence of different fillers and sbs modified bituminous blends on fatigue, self-healing and thixotropic performance of mastics. *Road Materials and Pavement Design*, 20(3):656–670.
- [McCulloch and Pitts, 1943] McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5:115–133.
- [Ministry of Road Transport and Highways, 2013] Ministry of Road Transport and Highways (2013). *Specifications for Road and Bridges Works, 5th Revision*. New Delhi, India.
- [Miró et al., 2017] Miró, R., Martínez, A. H., Pérez-Jiménez, F. E., Botella, R., and Álvarez, A. (2017). Effect of filler nature and content on the bituminous mastic behaviour under cyclic loads. *Construction and Building Materials*, 132:33–42.
- [Mirzahosseini et al., 2011] Mirzahosseini, M. R., Aghaeifar, A., Alavi, A. H., Gandomi, A. H., and Seyednour, R. (2011). Permanent deformation analysis of asphalt mixtures using soft computing techniques. *Expert Systems with Applications*, 38(5):6081–6100.

- [Mockus et al., 1978] Mockus, J., Tiesis, V., and Zilinskas, A. (1978). The application of bayesian methods for seeking the extremum. *Towards global optimization*, 2(117-129):2.
- [Modarres and Alinia Bengar, 2019] Modarres, A. and Alinia Bengar, P. (2019). Investigating the indirect tensile stiffness, toughness and fatigue life of hot mix asphalt containing copper slag powder. *International Journal of Pavement Engineering*, 20(8):977–985.
- [Montoya and Haddock, 2019] Montoya, M. A. and Haddock, J. E. (2019). Estimating asphalt mixture volumetric properties using seemingly unrelated regression equations approaches. *Construction and Building Materials*, 225:829–837.
- [Morgan and Sonquist, 1963] Morgan, J. N. and Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American statistical association*, 58(302):415–434.
- [Nemati and Dave, 2018] Nemati, R. and Dave, E. V. (2018). Nominal property based predictive models for asphalt mixture complex modulus (dynamic modulus and phase angle). *Construction and Building Materials*, 158:308–319.
- [Ng, 2004] Ng, A. Y. (2004). Feature selection, l_1 vs. l_2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on Machine learning*, page 78.
- [Nguyen et al., 2012] Nguyen, H. T. T., Luong, H. N., and Ahn, C. W. (2012). An entropy approach to evaluation relaxation for bayesian optimization algorithm. *International Journal of Innovative Computing, Information and Control*, 8:6371–6388.
- [Nikolaides, 2015] Nikolaides, A. (2015). *Pavements, Materials and Control of Quality*. CRC Press, Taylor & Francis Group.
- [Nilsson, 1998a] Nilsson, N. (1998a). *Artificial Intelligence: A New Synthesis*. The Morgan Kaufmann Series in Artificial Intelligence. Elsevier Science.
- [Nilsson, 1998b] Nilsson, N. (1998b). *Introduction to Machine Learning*. MIT Press (to appear).
- [Ogundipe, 2016] Ogundipe, O. M. (2016). Marshall stability and flow of lime-modified asphalt concrete. *Transportation Research Procedia*, 14:685–693.
- [Oh et al., 2020] Oh, C., Han, S., and Jeong, J. (2020). Time-series data augmentation based on interpolation. *Procedia Computer Science*, 175:64–71.
- [Olard and Di Benedetto, 2003] Olard, F. and Di Benedetto, H. (2003). General “2s2p1d” model and relation between the linear viscoelastic behaviours of bituminous binders and mixes. *Road materials and pavement design*, 4(2):185–224.
- [Ozgan, 2011] Ozgan, E. (2011). Artificial neural network based modelling of the marshall stability of asphalt concrete. *Expert Systems with Applications*, 38(5):6025–6030.
- [Ozsahin and Oruc, 2008] Ozsahin, T. S. and Oruc, S. (2008). Neural network model for resilient modulus of emulsified asphalt mixtures. *Construction and Building Materials*, 22(7):1436–1445.

- [Padhi et al., 2019] Padhi, S., Millstein, T., Nori, A., and Sharma, R. (2019). Overfitting in synthesis: Theory and practice. In *International Conference on Computer Aided Verification*, pages 315–334. Springer.
- [Pallant, 2020] Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. McGraw-hill education (UK).
- [Paris et al., 2003] Paris, G., Robilliard, D., and Fonlupt, C. (2003). Exploring overfitting in genetic programming. In *International Conference on Artificial Evolution (Evolution Artificielle)*, pages 267–277. Springer.
- [Pasandín and Pérez, 2015] Pasandín, A. and Pérez, I. (2015). Overview of bituminous mixtures made with recycled concrete aggregates. *Construction and Building Materials*, 74:151–161.
- [Pasetto and Baldo, 2015] Pasetto, M. and Baldo, N. (2015). Computational analysis of the creep behaviour of bituminous mixtures. *Construction and Building Materials*, 94:784–790.
- [Pasetto and Baldo, 2016] Pasetto, M. and Baldo, N. (2016). Numerical visco-elastoplastic constitutive modelization of creep recovery tests on hot mix asphalt. *Journal of Traffic and Transportation Engineering (English Edition)*, 3(5):390–397.
- [Phinyomark et al., 2013] Phinyomark, A., Quaine, F., Charbonnier, S., Serviere, C., Tarpin-Bernard, F., and Laurillau, Y. (2013). Emg feature evaluation for improving myoelectric pattern recognition robustness. *Expert Systems with applications*, 40(12):4832–4840.
- [Phung et al., 2023] Phung, B.-N., Le, T.-H., Nguyen, T.-A., Hoang, H.-G. T., and Ly, H.-B. (2023). Novel approaches to predict the marshall parameters of basalt fiber asphalt concrete. *Construction and Building Materials*, 400:132847.
- [Poh et al., 2018] Poh, C. Q., Ubeynarayana, C. U., and Goh, Y. M. (2018). Safety leading indicators for construction sites: A machine learning approach. *Automation in construction*, 93:375–386.
- [Poole et al., 1998] Poole, D., Mackworth, A., and Goebel, R. (1998). *Computational Intelligence: A Logical Approach*. Oxford University Press.
- [Prechelt, 2002] Prechelt, L. (2002). Early stopping-but when? In *Neural Networks: Tricks of the trade*, pages 55–69. Springer.
- [Prokhorenkova et al., 2018] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A. (2018). Catboost: unbiased boosting with categorical features. *Advances in neural information processing systems*, 31.
- [Quinlan, 1986] Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1:81–106.
- [Quinlan, 1993] Quinlan, J. R. (1993). C 4.5: Programs for machine learning. *The Morgan Kaufmann Series in Machine Learning*.

- [Rahman et al., 2019] Rahman, A. A., Mendez Larrain, M. M., and Tarefder, R. A. (2019). Development of a nonlinear rutting model for asphalt concrete based on weibull parameters. *International journal of pavement engineering*, 20(9):1055–1064.
- [Raposeiras et al., 2018] Raposeiras, A., Movilla-Quesada, D., Bilbao-Novoa, R., Cifuentes, C., Ferrer-Norambuena, G., and Castro-Fresno, D. (2018). The use of copper slags as an aggregate replacement in asphalt mixes with rap: Physical–chemical and mechanical behavioural analysis. *Construction and Building Materials*, 190:427–438.
- [Rashwan, 2016] Rashwan, N. (2016). Hot mix asphalt (hma) performance as affected by limestone powder filler content. *World Applied Sciences Journal*, 34(2):237–244.
- [Rasmussen, 2003] Rasmussen, C. E. (2003). Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer.
- [Rich and Knight, 1991] Rich, E. and Knight, K. (1991). *Artificial Intelligence*. Artificial Intelligence Series. McGraw-Hill.
- [Rix et al., 1995] Rix, G. J., Baker, N. C., Jacobs, L. J., Vanegas, J., and Zureick, A. H. (1995). Infrastructure assessment, rehabilitation, and reconstruction. In *Proceedings Frontiers in Education 1995 25th Annual Conference. Engineering Education for the 21st Century*, volume 2, pages 4c1–11. IEEE.
- [Rondinella et al., 2023] Rondinella, F., Oreto, C., Abbondati, F., and Baldo, N. (2023). Laboratory investigation and machine learning modeling of road pavement asphalt mixtures prepared with construction and demolition waste and rap. *Sustainability*, 15(23):16337.
- [Rumelhart et al., 1986] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088):533–536.
- [Russell and Norvig, 2010] Russell, S. and Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. Prentice Hall, 3 edition.
- [Sakhaeifar et al., 2015] Sakhaeifar, M. S., Kim, Y. R., and Kabir, P. (2015). New predictive models for the dynamic modulus of hot mix asphalt. *Construction and Building Materials*, 76:221–231.
- [Saleh and Fleyeh, 2022] Saleh, R. and Fleyeh, H. (2022). Using supervised machine learning to predict the status of road signs. *Transportation research procedia*, 62:221–228.
- [Santagata et al., 2009] Santagata, F. A., Canestrari, F., and Pasquini, E. (2009). Controllo della produzione, verifica delle prestazioni e predisposizione di norme tecniche di capitolato per miscele di conglomerato bituminoso tipo asphalt rubber: Rapporto finale.
- [Sarkar et al., 2018] Sarkar, D., Bali, R., and Sharma, T. (2018). Practical machine learning with python. *A Problem-Solvers Guide To Building Real-World Intelligent Systems*. Berkely: Apress.
- [Sarkar et al., 2016] Sarkar, D., Pal, M., Sarkar, A. K., and Mishra, U. (2016). Evaluation of the properties of bituminous concrete prepared from brick-stone mix aggregate. *Advances in Materials Science and Engineering*, 2016.

- [Schneider et al., 2011] Schneider, M., Romer, M., Tschudin, M., and Bolio, H. (2011). Sustainable cement production—present and future. *Cement and concrete research*, 41(7):642–650.
- [Schwartz et al., 1991] Schwartz, J. H., Jessell, T. M., and Kandel, E. R. (1991). *Principles of neural science*. Elsevier New York.
- [Schweikert et al., 2014] Schweikert, A., Chinowsky, P., Espinet, X., and Tarbert, M. (2014). Climate change and infrastructure impacts: Comparing the impact on roads in ten countries through 2100. *Procedia Engineering*, 78:306–316.
- [Scimemi et al., 2016] Scimemi, G. F., Turetta, T., and Celauro, C. (2016). Backcalculation of airport pavement moduli and thickness using the lévy ant colony optimization algorithm. *Construction and Building Materials*, 119:288–295.
- [Sebaaly et al., 1986] Sebaaly, B. E., Mamlouk, M. S., and Davies, T. G. (1986). Dynamic analysis of falling weight deflectometer data. *Transportation Research Record*, 1070:63–68.
- [Senadjki et al., 2022] Senadjki, A., Awal, I. M., Nee, A. Y. H., and Ogbeibu, S. (2022). The belt and road initiative (bri): A mechanism to achieve the ninth sustainable development goal (sdg). *Journal of Cleaner Production*, 372:133590.
- [Shahriari et al., 2015] Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and De Freitas, N. (2015). Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175.
- [Shapley et al., 1953] Shapley, L. S. et al. (1953). A value for n-person games. *Contributions to the Theory of Games*.
- [Shepherd, 2003] Shepherd, G. M. (2003). *The synaptic organization of the brain*. Oxford university press.
- [Sholevar et al., 2022] Sholevar, N., Golroo, A., and Esfahani, S. R. (2022). Machine learning techniques for pavement condition evaluation. *Automation in Construction*, 136:104190.
- [Shukla and Das, 2008] Shukla, P. K. and Das, A. (2008). A re-visit to the development of fatigue and rutting equations used for asphalt pavement design. *International Journal of Pavement Engineering*, 9(5):355–364.
- [Siddharthan et al., 1991] Siddharthan, R., Norris, G. M., and Epps, J. A. (1991). Use of fwd data for pavement material characterization and performance. *Journal of transportation engineering*, 117(6):660–678.
- [Silva et al., 2016] Silva, I. N., Spatti, D. H., Flauzino, R. A., Liboni, L., and Alves, S. (2016). *Artificial neural networks: a practical course*. Springer.
- [Snoek et al., 2012] Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. *arXiv preprint arXiv:1206.2944*.

- [Sousa et al., 1995] Sousa, J. B., Way, G., Harvey, J. T., and Hines, M. (1995). Comparison of mix design concepts. *Transportation Research Record*, 1492:151.
- [Specht et al., 2007] Specht, L. P., Khatchatourian, O., Brito, L. A. T., and Ceratti, J. A. P. (2007). Modeling of asphalt-rubber rotational viscosity by statistical analysis and neural networks. *Materials Research*, 10:69–74.
- [Srinivas et al., 2009] Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M. (2009). Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*.
- [Subasi, 2020] Subasi, A. (2020). *Practical machine learning for data analysis using python*. Academic Press.
- [Subramaniam et al., 2023] Subramaniam, N., Akbar, S., Situ, H., Ji, S., and Parikh, N. (2023). Sustainable development goal reporting: Contrasting effects of institutional and organisational factors. *Journal of Cleaner Production*, 411:137339.
- [Sun et al., 2022] Sun, B., Hao, P., Zhang, H., and Liu, J. (2022). Establishment and verification of a developed viscoelastic damage model and creep instability criterion for modified fine asphalt mortar. *Materials and Structures*, 55(3):107.
- [Svilar et al., 2022] Svilar, M., Peško, I., and Šešlija, M. (2022). Model for estimating the modulus of elasticity of asphalt layers using machine learning. *Applied Sciences*, 12(20):10536.
- [Talvik and Aavik, 2009] Talvik, O. and Aavik, A. (2009). Use of fwd deflection basin parameters (sci, bdi, bci) for pavement condition assessment. *Transformation*, 7(870):260.
- [Tapkın et al., 2009] Tapkın, S., Çevik, A., and Uşar, Ü. (2009). Accumulated strain prediction of polypropylene modified marshall specimens in repeated creep test using artificial neural networks. *Expert Systems with Applications*, 36(8):11186–11197.
- [Tapkın et al., 2010] Tapkın, S., Çevik, A., and Uşar, Ü. (2010). Prediction of marshall test results for polypropylene modified dense bituminous mixtures using neural networks. *Expert Systems with Applications*, 37(6):4660–4670.
- [Tarefder et al., 2005] Tarefder, R. A., White, L., and Zaman, M. (2005). Neural network model for asphalt concrete permeability. *Journal of Materials in Civil Engineering*, 17(1):19–27.
- [Tessema et al., 2022] Tessema, A. T., Wolelaw, N. M., Alene, G. A., et al. (2022). Experimental evaluation of coffee husk ash as a filler in hot mix asphalt concrete productions. *Advances in Civil Engineering*, 2022.
- [Thompson and Garg, 1997] Thompson, M. and Garg, N. (1997). Mechanistic-empirical evaluation of the mn/road low volume road test sections. *Project IHR-535. University of Illinois at Urbana-Champaign*, 0.2:0–4.

- [Tikhonov, 1963] Tikhonov, A. N. (1963). On the solution of ill-posed problems and the method of regularization. In *Doklady Akademii Nauk*, volume 151, pages 501–504. Russian Academy of Sciences.
- [Tiwari and Satyam, 2021] Tiwari, N. and Satyam, N. (2021). Evaluation of strength and water susceptibility performance of polypropylene fiber–reinforced and silica fume–modified hot mix asphalt. *Advances in Civil Engineering Materials*, 10(1):380–395.
- [Turing, 1950] Turing, A. M. (1950). *Mind*. *Mind*, 59(236):433–460.
- [Ullidtz, 1987] Ullidtz, P. (1987). *Pavement Analysis. Developments in Civil Engineering*, 19. Elsevier.
- [Ullidtz, 2000] Ullidtz, P. (2000). Will nonlinear backcalculation help? In *Nondestructive Testing of Pavements and Backcalculation of Moduli*, pages 14–22. American Society for Testing and Materials.
- [Ullidtz and Coetzee, 1995] Ullidtz, P. and Coetzee, N. (1995). Analytical procedures in nondestructive testing pavement evaluation. *Transportation research record*, 1482:61–66.
- [Underwood and Kim, 2011] Underwood, S. B. and Kim, R. Y. (2011). Viscoelastoplastic continuum damage model for asphalt concrete in tension. *Journal of engineering mechanics*, 137(11):732–739.
- [UNI EN 1097-2, 2020] UNI EN 1097-2 (2020). *Tests for Mechanical and Physical Properties of Aggregates—Part 2: Methods for the Determination of Resistance to Fragmentation*. Brussels, Belgium.
- [UNI EN 1097-8, 2020] UNI EN 1097-8 (2020). *Tests for Mechanical and Physical Properties of Aggregates-Part 8: Determination of the Polished Stone Value*. Brussels, Belgium.
- [UNI EN 12592, 2007] UNI EN 12592 (2007). *Bitumen and Bituminous Binders - Determination of Solubility*. Brussels, Belgium.
- [UNI EN 12593, 2015] UNI EN 12593 (2015). *Bitumen and Bituminous Binders - Determination of the Fraass Breaking Point*. Brussels, Belgium.
- [UNI EN 12697-12, 2018] UNI EN 12697-12 (2018). *Bituminous Mixtures—Test Methods—Part 12: Determination of the Water Sensitivity of Bituminous Specimens*. Brussels, Belgium.
- [UNI EN 12697-23, 2018] UNI EN 12697-23 (2018). *Determination of the Indirect Tensile Strength of Bituminous Specimens*. Brussels, Belgium.
- [UNI EN 12697-26, 2018] UNI EN 12697-26 (2018). *Bituminous mixtures - Test methods - Part 26: Stiffness*. Brussels, Belgium.
- [UNI EN 12697-30, 2018] UNI EN 12697-30 (2018). *Bituminous mixtures - Test methods - Part 30: Specimen preparation by impact compactor*. Brussels, Belgium.

BIBLIOGRAFIA

- [UNI EN 12697-31, 2019] UNI EN 12697-31 (2019). *Bituminous mixtures - Test methods - Part 31: Specimen preparation by gyratory compactor*. Brussels, Belgium.
- [UNI EN 12697-32, 2019] UNI EN 12697-32 (2019). *Bituminous mixtures - Test methods - Part 32: Specimen preparation by vibratory compactor*. Brussels, Belgium.
- [UNI EN 12697-33, 2019] UNI EN 12697-33 (2019). *Bituminous mixtures - Test methods - Part 33: Specimen prepared by roller compactor*. Brussels, Belgium.
- [UNI EN 12697-34, 2020] UNI EN 12697-34 (2020). *Bituminous mixtures - Test methods - Part 34: Marshall testd*. Brussels, Belgium.
- [UNI EN 12697-8, 2019] UNI EN 12697-8 (2019). *Bituminous Mixtures—Test Methods—Part 8: Determination of Void Characteristics of Bituminous Specimens*. Brussels, Belgium.
- [UNI EN 13108-1, 2016] UNI EN 13108-1 (2016). *Bituminous Mixtures—Material Specifications—Part 1: Asphalt Concrete*. Brussels, Belgium.
- [UNI EN 13108-20, 2016] UNI EN 13108-20 (2016). *Bituminous Mixtures—Material Specifications—Part 20: Type Testing*. Brussels, Belgium.
- [UNI EN 13398, 2017] UNI EN 13398 (2017). *Bitumen and Bituminous Binders - Determination of the Elastic Recovery of Modified Bitumen*. Brussels, Belgium.
- [UNI EN 13589, 2018] UNI EN 13589 (2018). *Bitumen and Bituminous Binders - Determination of the Tensile Properties of Modified Bitumen by the Force Ductility Method*. Brussels, Belgium.
- [UNI EN 1426, 2015] UNI EN 1426 (2015). *Bitumen and Bituminous Binders - Determination of Needle Penetration*. Brussels, Belgium.
- [UNI EN 1427, 2015] UNI EN 1427 (2015). *Bitumen and Bituminous Binders - Determination of the Softening Point - Ring and Ball Method*. Brussels, Belgium.
- [UNI EN 933-3, 2012] UNI EN 933-3 (2012). *Tests for Geometrical Properties of Aggregates—Part 3: Determination of Particle Shape — Flakiness Index*. Brussels, Belgium.
- [UNI EN 933-4, 2008] UNI EN 933-4 (2008). *Tests for Geometrical Properties of Aggregates—Part 4: Determination of Particle Shape — Shape Index*. Brussels, Belgium.
- [UNI EN 933-8, 2016] UNI EN 933-8 (2016). *Tests for Geometrical Properties of Aggregates—Part 8: Assessment of Fines — Sand Equivalent Test*. Brussels, Belgium.
- [UNI EN 933-9, 2022] UNI EN 933-9 (2022). *Tests for Geometrical Properties of Aggregates-Part 9: Assessment of Fines - Methylene Blue Test*. Brussels, Belgium.
- [UNI EN ISO 2592, 2017] UNI EN ISO 2592 (2017). *Petroleum Products — Determination of Flash and Fire Points — Cleveland Open Cup Method*. Brussels, Belgium.

BIBLIOGRAFIA

- [Upadhya et al., 2023] Upadhya, A., Thakur, M., Sihag, P., Kumar, R., Kumar, S., Afeeza, A., Afzal, A., and Saleel, C. A. (2023). Modelling and prediction of binder content using latest intelligent machine learning algorithms in carbon fiber reinforced asphalt concrete. *Alexandria Engineering Journal*, 65:131–149.
- [USDOT, 2011] USDOT, F. (2011). Use of nondestructive testing in the evaluation of airport pavements. *Advisory Circular 150/5370-11B*.
- [Uzan, 1994] Uzan, J. (1994). Advanced backcalculation techniques. In *Nondestructive Testing of Pavements and Backcalculation of Moduli: Second Volume*, pages 3–37. ASTM International.
- [Van Thanh and Feng, 2013] Van Thanh, D. and Feng, C. P. (2013). Study on marshall and rutting test of sma at abnormally high temperature. *Construction and Building Materials*, 47:1337–1341.
- [Vapnik, 1991] Vapnik, V. (1991). Principles of risk minimization for learning theory. *Advances in neural information processing systems*, 4.
- [Veytskin et al., 2015] Veytskin, Y., Bobko, C., Castorena, C., and Kim, Y. R. (2015). Nanoindentation investigation of asphalt binder and mastic cohesion. *Construction and Building Materials*, 100:163–171.
- [Wang et al., 2011] Wang, H., Al-Qadi, I. L., Faheem, A. F., Bahia, H. U., Yang, S.-H., and Reinke, G. H. (2011). Effect of mineral filler characteristics on asphalt mastic and mixture rutting potential. *Transportation Research Record*, 2208(1):33–39.
- [Wang et al., 2021] Wang, H., Xie, P., Ji, R., and Gagnon, J. (2021). Prediction of airfield pavement responses from surface deflections: comparison between the traditional backcalculation approach and the ann model. *Road Materials and Pavement Design*, 22(9):1930–1945.
- [Wang et al., 2017] Wang, L., Gong, H., Hou, Y., Shu, X., and Huang, B. (2017). Advances in pavement materials, design, characterisation, and simulation. *Road Materials and Pavement Design*, 18(sup3):1–11.
- [Weil et al., 1992] Weil, G. J. et al. (1992). Non-destructive testing of bridge, highway and airport pavements. In *Fourth International Conference on Ground Penetrating Radar*, pages 259–266. European Association of Geoscientists & Engineers.
- [Winston, 1992] Winston, P. (1992). *Artificial Intelligence*. A-W Series in Computerscience. Addison-Wesley Publishing Company.
- [Witczak et al., 2007] Witczak, M., El-Basyouny, M., and El-Badawy, S. (2007). Incorporation of the new (2005) e* predictive model in the mepdg. *NCHRP 1-40D Final Report*.
- [Witten and James, 2013] Witten, D. and James, G. (2013). *An introduction to statistical learning with applications in R*. springer publication.

- [Wołczowski and Zdunek, 2017] Wołczowski, A. and Zdunek, R. (2017). Electromyography and mechanomyography signal recognition: Experimental analysis using multi-way array decomposition methods. *Biocybernetics and Biomedical Engineering*, 37(1):103–113.
- [Wu and Montalvo, 2021] Wu, S. and Montalvo, L. (2021). Repurposing waste plastics into cleaner asphalt pavement materials: A critical literature review. *Journal of Cleaner Production*, 280:124355.
- [Xiao et al., 2009] Xiao, F., Amirkhani, S., and Juang, C. H. (2009). Prediction of fatigue life of rubberized asphalt concrete mixtures containing reclaimed asphalt pavement using artificial neural networks. *Journal of Materials in Civil Engineering*, 21(6):253–261.
- [Xiao and Amirkhani, 2009] Xiao, F. and Amirkhani, S. N. (2009). Artificial neural network approach to estimating stiffness behavior of rubberized asphalt concrete containing reclaimed asphalt pavement. *Journal of Transportation Engineering*, 135(8):580–589.
- [Xu et al., 2016] Xu, B., Chen, J., Zhou, C., and Wang, W. (2016). Study on marshall design parameters of porous asphalt mixture using limestone as coarse aggregate. *Construction and Building Materials*, 124:846–854.
- [Xu et al., 2002] Xu, B., Ranjithan, S. R., and Kim, Y. R. (2002). New relationships between falling weight deflectometer deflections and asphalt pavement layer condition indicators. *Transportation research record*, 1806(1):48–56.
- [Xue et al., 2009] Xue, Y., Hou, H., Zhu, S., and Zha, J. (2009). Utilization of municipal solid waste incineration ash in stone mastic asphalt mixture: Pavement performance and environmental impact. *Construction and Building Materials*, 23(2):989–996.
- [Yang et al., 2022] Yang, E., Yang, Q., Li, J., Zhang, H., Di, H., and Qiu, Y. (2022). Establishment of icing prediction model of asphalt pavement based on support vector regression algorithm and bayesian optimization. *Construction and Building Materials*, 351:128955.
- [Yilmaz et al., 2011] Yilmaz, M., Kök, B. V., and Kuloğlu, N. (2011). Effects of using asphaltite as filler on mechanical properties of hot mix asphalt. *Construction and Building Materials*, 25(11):4279–4286.
- [Ying, 2019] Ying, X. (2019). An overview of overfitting and its solutions. In *Journal of physics: Conference series*, volume 1168, page 022022. IOP Publishing.
- [Yun and Kim, 2013] Yun, T. and Kim, Y. R. (2013). Viscoelastoplastic modeling of the behavior of hot mix asphalt in compression. *KSCE Journal of Civil Engineering*, 17(6):1323–1332.
- [Zaumanis et al., 2016] Zaumanis, M., Mallick, R. B., and Frank, R. (2016). 100% hot mix asphalt recycling: challenges and benefits. *Transportation research procedia*, 14:3493–3502.
- [Zavrtanik et al., 2016] Zavrtanik, N., Prosen, J., Tušar, M., and Turk, G. (2016). The use of artificial neural networks for modeling air void content in aggregate mixture. *Automation in Construction*, 63:155–161.

- [Zeiada et al., 2020] Zeiada, W., Dabous, S. A., Hamad, K., Al-Ruzouq, R., and Khalil, M. A. (2020). Machine learning for pavement performance modelling in warm climate regions. *Arabian journal for science and engineering*, 45(5):4091–4109.
- [Zhang, 2019] Zhang, H. (2019). Introductory chapter: Asphalt and asphalt mixture. *Asphalt and Asphalt Mixtures*, page 3.
- [Zhang et al., 2019] Zhang, M. R., Lucas, J., Hinton, G., and Ba, J. (2019). Lookahead optimizer: k steps forward, 1 step back. *arXiv preprint arXiv:1907.08610*.
- [Zhang et al., 2022] Zhang, R., Sias, J. E., and Dave, E. V. (2022). Using mix design information for modelling of fundamental viscoelasticity of asphalt mixtures. *Construction and Building Materials*, 329:127029.
- [Zhao and Yang, 2023] Zhao, W. and Yang, Q. (2023). Design and performance evaluation of a new green pavement: 100% recycled asphalt pavement and 100% industrial solid waste. *Journal of Cleaner Production*, 421:138483.
- [Zheng et al., 2021] Zheng, M., Chen, W., Ding, X., Zhang, W., and Yu, S. (2021). Comprehensive life cycle environmental assessment of preventive maintenance techniques for asphalt pavement. *Sustainability*, 13(9):4887.
- [Zhou et al., 2004] Zhou, F., Scullion, T., and Sun, L. (2004). Verification and modeling of three-stage permanent deformation behavior of asphalt mixes. *Journal of Transportation Engineering*, 130(4):486–494.
- [Ziari et al., 2018] Ziari, H., Amini, A., Goli, A., and Mirzaiyan, D. (2018). Predicting rutting performance of carbon nano tube (cnt) asphalt binders using regression models and neural networks. *Construction and Building Materials*, 160:415–426.
- [Zoorob and Suparma, 2000] Zoorob, S. and Suparma, L. (2000). Laboratory design and investigation of the properties of continuously graded asphaltic concrete containing recycled plastics aggregate replacement (plastiphalt). *Cement and concrete composites*, 22(4):233–242.
- [Zou et al., 2020] Zou, L., Yi, L., and Sato, M. (2020). On the use of lateral wave for the interlayer debonding detecting in an asphalt airport pavement using a multistatic gpr system. *IEEE Transactions on Geoscience and Remote Sensing*, 58(6):4215–4224.
- [Zumrawi and Khalill, 2017] Zumrawi, M. M. and Khalill, F. O. (2017). Experimental study of steel slag used as aggregate in asphalt mixture. *American Journal of Construction and Building Materials*, 2(1):26–32.

Appendice A

Lista delle Pubblicazioni

In questa sezione vengono brevemente riportati i lavori scientifici pubblicati in ambito nazionale ed internazionale durante il periodo di dottorato. I contenuti ed i risultati sperimentali pubblicati in alcuni di questi lavori sono stati descritti nei capitoli precedenti.

1. Baldo, Nicola; Miani, Matteo; **Rondinella, Fabio**; Celauro, Clara. Machine Learning Approach to Determine Airport Asphalt Concrete Layer Moduli using Heavy Weight Deflectometer Data. *Sustainability*, 2021. Special Issue: Transportation Safety and Pavement Management. "Editor's Choice in recognition of some of the most exciting work published in the various research areas of the journal."
<https://doi.org/10.3390/su13168831>
2. Baldo, Nicola; Miani, Matteo; **Rondinella, Fabio**; Pasetto, Marco. Recycling of Waste Materials Using Bitumen Emulsion for Road Pavement Stabilized Base Courses: A Laboratory Investigation. *IOP Conference Series: Materials Science and Engineering*, 6th World Multidisciplinary Civil Engineering - Architecture - Urban Planning Symposium - WMCAUS 30 August - 3 September 2021, Prague, Czech Republic. <http://dx.doi.org/10.1088/1757-899x/1203/2/022111>
3. Baldo, Nicola; Miani, Matteo; **Rondinella, Fabio**; Vacková, Pavla; Valentin, Jan. Performance Prediction of Fine-Grained Asphalt Concretes with Different Quarry Fillers by Machine Learning Approaches. *IOP Conference Series: Materials Science and Engineering*, 6th World Multidisciplinary Civil Engineering - Architecture - Urban Planning Symposium - WMCAUS 30 August - 3 September 2021, Prague, Czech Republic. <http://dx.doi.org/10.1088/1757-899x/1203/2/022113>
4. Baldo, Nicola; Miani, Matteo; **Rondinella, Fabio**; Celauro, Clara. Artificial Neural Network Prediction of Airport Pavement Moduli Using Interpolated Surface Deflection Data. *IOP Conference Series: Materials Science and Engineering*, 6th World Multidisciplinary Civil Engineering - Architecture - Urban Planning Symposium - WMCAUS 30 August - 3 September 2021, Prague, Czech Republic.
<http://dx.doi.org/10.1088/1757-899x/1203/2/022112>

5. Miani, Matteo; Dunnhofer, Matteo; **Rondinella, Fabio**; Manthos, Evangelos; Valentin, Jan; Micheloni, Christian; Baldo, Nicola. Bituminous Mixtures Experimental Data Modeling Using a Hyperparameters Optimized Machine Learning Approach. *Applied Sciences*, 2021. <https://doi.org/10.3390/app112411710>
6. Baldo, Nicola; Miani, Matteo; **Rondinella, Fabio**; Valentin, Jan; Vackcová, Pavla; Manthos, Evangelos. Stiffness Data of High-Modulus Asphalt Concretes for Road Pavements: Predictive Modeling by Machine - Learning. *Coatings*, 2022. "Feature paper in recognition of the most advanced research with significant potential for high impact in the field." <https://doi.org/10.3390/coatings12010054>
7. Baldo, Nicola; **Rondinella, Fabio**; Daneluz, Fabiola; Pasetto, Marco. Foamed Bitumen Mixtures for Road Construction Made with 100% Waste Materials: A Laboratory Study. *Sustainability*, 2022. <https://doi.org/10.3390/su14106056>
8. Baldo, Nicola; Miani, Matteo; **Rondinella, Fabio**; Manthos, Evangelos; Valentin, Jan. Road Pavement Asphalt Concretes for Thin Wearing Layers: A Machine Learning Approach towards Stiffness Modulus and Volumetric Properties Prediction. *Periodica Polytechnica Civil Engineering*, 2022. <https://doi.org/10.3311/PPci.19996>
9. **Rondinella, Fabio**; Baldo, Nicola; Celauro, Clara. Improved Moduli Estimation of Airport Asphalt Concrete Layer Using a Machine Learning Approach Based on Heavy Weight Deflectometer Data. *Sustainable Pavements and Road Materials: Proceedings of the 7th SIV Arena*, 5 September - 9 September 2022, Naples, Italy.
10. Baldo, Nicola; Miani, Matteo; **Rondinella, Fabio**; Manthos, Evangelos Valentin, Jan. A Machine Learning Approach for the Prediction of Very Thin Wearing Layers Asphalt Concretes Volumetric Properties and Performance. *IOP Conference Series: Materials Science and Engineering*, 7th World Multidisciplinary Civil Engineering - Architecture - Urban Planning Symposium – WMCAUS 5 September - 9 September 2022, Prague, Czech Republic.
11. Tiwari, Nitin; **Rondinella, Fabio**; Satyam, Neelima; Baldo, Nicola. Silica Fume As A Surrogate Filler In Asphalt Concrete Mixtures: Laboratory Investigation And a Machine Learning-Based Prediction. *IOP Conference Series: Materials Science and Engineering*, 7th World Multidisciplinary Civil Engineering - Architecture - Urban Planning Symposium – WMCAUS 5 September - 9 September 2022, Prague, Czech Republic.
12. Baldo, Nicola; **Rondinella, Fabio**; Celauro, Clara. Prediction of Airport Pavement Moduli by Machine Learning Methodology Using Non - Destructive Field Testing Data Augmentation. *International Conference on Trends on Construction in the Post-Digital Era*. 6-9 September 2023, Guimarães, Portugal.
13. Baldo, Nicola; Indri, Valentina; **Rondinella, Fabio**; Daneluz, Fabiola. Multivariate Regression of Road Segments' Accident Data in Italia Rural Networks. *AIIT 3rd International Conference*. New Scenarios for Transport Infrastructure and Systems Transition to Inclusivity, Resilience, and Sustainability. 15 - 16 September 2022, Rome, Italy.

14. Tiwari, Nitin; **Rondinella, Fabio**; Satyam, Neelima; Baldo, Nicola. Alternative Fillers in Asphalt Concrete Mixtures: Laboratory Investigation and Machine Learning Modeling towards Mechanical Performance Prediction. *Materials*, 2023. <https://doi.org/10.3390/ma16020807>
15. **Rondinella, Fabio**; Daneluz, Fabiola; Vackcová, Pavla; Valentin, Jan; Baldo, Nicola; Volumetric Properties and Stiffness Modulus of Asphalt Concrete Mixtures made with Selected Quarry Fillers: Experimental Investigation and Machine Learning Prediction. *Materials*, 2023. <https://doi.org/10.3390/ma16031017>
16. Baldo, Nicola; Indri, Valentina; **Rondinella, Fabio**; Daneluz, Fabiola. Multivariate Regression of Road Segments' Accident Data in Italia Rural Networks. *European Transport*, 2023. <https://dx.doi.org/10.48295/ET.2023.91.6>
17. Tiwari, Nitin; **Rondinella, Fabio**; Satyam, Neelima; Baldo, Nicola. Experimental and Machine Learning Approach to Investigate the Mechanical Performance of Asphalt Mixtures with Silica Fume Filler. *Applied Sciences*, 2023. <https://doi.org/10.3390/app13116664>
18. Baldo, Nicola; **Rondinella, Fabio**; Valentin, Jan; Król, Jan B.; Gajewski, Marcin D. An Artificial Neural Network Approach to Predict Asphalt Mixtures' Stiffness Modulus Based on Testing Frequency and Temperature. *IOP Conference Series: Materials Science and Engineering*, 8th World Multidisciplinary Civil Engineering - Architecture - Urban Planning Symposium – WMCAUS 4 September - 8 September 2023, Prague, Czech Republic.
19. **Rondinella, Fabio**; Daneluz, Fabiola; Hofko, Bernhard; Baldo, Nicola; Improved Predictions of Asphalt Concretes' Dynamic Modulus and Phase Angle Using Decision-Tree Based Categorical Boosting Model. *Construction and Building Materials*, 2023. <https://doi.org/10.1016/j.conbuildmat.2023.132709>
20. **Rondinella, Fabio**; Daneluz, Fabiola; Hofko, Bernhard; Baldo, Nicola. A Machine Learning Approach for the Simultaneous Prediction of Dynamic Modulus and Phase Angle of Asphalt Concrete Mixtures. *International Conference on Advanced Research in Technologies, Information, Innovation and Sustainability - ARTIIS 18 October - 20 October 2023, Madrid, Spain*.
21. Baldo, Nicola; **Rondinella, Fabio**; Daneluz, Fabiola; Vacková, Pavla; Valentin, Jan; Gajewski, Marcin D.; Król, Jan B. Stiffness Moduli Modelling and Prediction in Four-Point Bending of Asphalt Mixtures: A Machine Learning-Based Framework. *CivilEng*, 2023. <https://doi.org/10.3390/civileng4040059>
22. **Rondinella, F.**; Oreto, Cristina; Abbondati, Francesco; Baldo, Nicola. Laboratory Investigation and Machine Learning Modeling of Road Pavement Asphalt Mixtures Prepared with Construction and Demolition Waste and RAP. *Sustainability*, 2023. <https://doi.org/10.3390/su152316337>

23. Baldo, N.; **Rondinella, Fabio**; Daneluz, Fabiola; Valentin, Jan; Vackcová, Pavla; Król, Jan B.; Gajewski, Marcin D. Asphalt mixtures' stiffness modulus prediction using a machine-learning approach based on temperature and frequency conditions. 8th International Conference on Bituminous Mixtures and Pavements - ICONFBMP 12 June - 14 June 2024, Thessaloniki, Greece.
24. Baldo, Nicola; **Rondinella, Fabio**; Daneluz, Fabiola; Valentin, Jan; Vackcová, Pavla; Manthos, Evangelos. Predicting volumetric properties and mechanical characteristics of asphalt concretes for thin wearing layers using a categorical boosting model. 8th International Conference on Bituminous Mixtures and Pavements - ICONFBMP 12 June - 14 June 2024, Thessaloniki, Greece.

Appendice B

Elenco delle Abbreviazioni

AC	Asphalt Concrete
AI	Artificial Intelligence
ANN	Artificial Neural Network
APMS	Airport Pavement Management System
AREA	AREA under deflection basin curve
AUPP	Area Under Pavement Profile
BC	Bitumen Content
BD	Brick Dust
BM	Bituminous Mixture
CART	Classification And Regression Tree
CCP	Cost Complexity Pruning
CL	Cantabro Loss
CV	Cross Validation
DBP	Deflection Basin Parameter
DNN	Deep Neural Network
DR	Deflection Ratio
DT	Decision Tree
ENP	Effective Number of Parameters
ESAL	Equivalent Single Axle Load
FA	Fly Ash
FI	Flakiness Index
FM	Fineness Modulus
FN	False Negative
FP	False Positive
FtB	Filler to Bitumen
FWD	Falling Weight Deflectometer
GB	Gradient Boosting
GBDT	Gradient Boosting Decision Tree
GP	Gaussian Process

APPENDICE B. ELENCO DELLE ABBREVIAZIONI

HS	Homogeneous Section
HWD	Heavy Weight Deflectometer
IG	Information Gain
IRI	International Roughness Index
ITS	Indirect Tensile Strength
ITSM	Indirect Tensile Stiffness Modulus
ITSR	Indirect Tensile Strength Ratio
KKT	Karush Kuhn Tucker
LA	Los Angeles
LD	Limestone Dust
LM	Levenberg Marquardt
LOOCV	Leave-One-Out Cross Validation
LR	Lasso Regression
M&R	Maintenance and Rehabilitation
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MBV	Methylene Blue Value
M-EPDG	Mechanistic-Empirical Pavement Design Guide
MF	Marshall Flow
ML	Machine Learning
MLE	MultiLayer Elastic
MLP	MultiLayer Perceptron
MoRTH	Indian Ministry of Road Transport and Highway
MQ	Marshall Quotient
MS	Marshall Stability
MSE	Mean Squared Error
NDT	Non Destructive Test
OB	Ordered Boosting
ODT	Oblivious Decision Tree
OPC	Ordinary Portland Cement
PG	Performance Grade
PSV	Polished Stone Value
R&B	Ring and Ball
RHA	Rice Husk Ash
RMSE	Root Mean Squared Error
ROME	ROad Moduli Evaluation
RR	Ridge Regression
RSS	Residual Sum of Squares
SBS	Styrene-Butadiene-Styrene
SCI	Surface Curvature Index
SDG	Sustainable Development Goal
SE	Sand Equivalent
SF	Silica Fume

APPENDICE B. ELENCO DELLE ABBREVIAZIONI

SHAP	SHapley Additive exPlanations
SHRP	Strategic Highway Research Program
SI	Shape Index
SMA	Stone Mastic Asphalt
SNN	Shallow Neural Network
SoA	State of Art
SVM	Support Vector Machine
SVR	Support Vector Regressor
TDZ	TouchDown Zone
TN	True Negative
TP	True Positive
TTSP	Time Temperature Superposition Principle
UCB	Upper Confidence Bound
VFB	Voids Filled with Bitumen
VMA	Voids in the Mineral Aggregate