SYM-OP-IS 2025
52nd Symposium on Operational Research
Palić, 7 - 10.9.2025.

UNIVERSITY OF BELGRADE
FACULTY OF ORGANIZATIONAL SCIENCES

Original article **https://doi.org/10.5281/zenodo.17533277**

DL-SPP: A DEEP LEARNING SYSTEM FOR SPARE PARTS PREDICTION IN FIELD SERVICE OPERATIONS

ALESSIO SCHIAVO[1,2], ALESSANDRO RENDA[3], PIETRO DUCANGE[4], FRANCESCO MARCELLONI[5]

[1] University of Pisa, Department of Information Engineering, Pisa, alessio.schiavo@phd.unipi.it, 0009-0005-2147-2853
[2] LogObject AG, Opfikon, Switzerland, alessio.schiavo@logobject.ch
[3] University of Trieste, Department of Engineering and Architecture, Trieste, alessandro.renda@dia.units.it 0000-0002-0482-5048
[4] University of Pisa, Department of Information Engineering, Pisa, pietro.ducange@unipi.it, 0000-0003-4510-1350
[5] University of Pisa, Department of Information Engineering, Pisa, francesco.marcelloni@unipi.it, 0000-0002-5895-876X

Abstract: *Accurate spare parts prediction is crucial in field service operations for the maintenance and repair of household appliances, where missing components frequently lead to repeated technician visits, increased costs, and diminished service quality. In this context, we present the DL-SPP (Deep Learning Spare Part Prediction) system that predicts the set of spare parts likely needed before the technician's intervention, using both structured appliance metadata and semantic representations of fault descriptions. The prediction task is framed as a multi-label classification problem, where the model learns to associate appliance types and fault reports with the components used in past repairs through a transformer-based architecture. Evaluated on a real-world dataset of repair tasks, DL-SPP achieves up to a 56% reduction in repeated field service technician visits, outperforming a state-of-the-art baseline that achieves 50%. These results highlight the DL-SPP system as a scalable and effective solution for enhancing repair planning and inventory optimization in operational field service contexts.*

Keywords: *Field Service Optimization, Spare Parts Prediction, Natural Language Processing, Repair Kit Problem, Deep Learning*

1. INTRODUCTION

Optimizing field service operations is critical for companies providing repair and maintenance of household appliances. When an appliance malfunctions, completing the repair during a single technician visit is essential to minimize operational costs and ensure customer satisfaction. Additional visits lead to increased labor, inventory, and travel expenses, along with diminished service quality. Consequently, accurately predicting the required spare parts prior to the intervention becomes a critical factor in improving operational efficiency.

Traditional approaches to spare parts planning have mainly addressed inventory management and demand forecasting, typically framing the problem as a combinatorial optimization task. In the context of the *Repair Kit Problem* (Smith et al., 1980), several studies have focused on identifying the optimal assortment of parts to be stocked in service vans, balancing holding costs with the penalties associated with incomplete repairs (Bijvank and Vis, 2010; Prak et al., 2017; Teunter, 2006). More recent contributions have explored heuristic and simulation-based methods for handling multiple-job routes, uncertain demands, or replenishment constraints (Pham and Gudrun P. Kiesmüller, 2022; Rippe and Gudrun P. Kiesmüller, 2023a; Rippe and Gudrun P Kiesmüller, 2023b). In these works, the probability that a given spare part *sp* is required for a specific repair task t_j is denoted as $p_{sp}^{t_j}$ and treated as a fixed input parameter. This probability is typically estimated using simple frequency counts from historical data. However, this assumption limits the flexibility and adaptability of such models in real-world scenarios, where the need for specific parts often depends on contextual and unstructured information such as fault descriptions.

Some studies (Saccani and Johansson, 2017) have proposed budget-aware repair kit strategies, incorporating sensor-based fault signals as Advance Demand Information (ADI), while others have integrated scheduling components to jointly optimize service routing and spare part planning (Pham and Gudrun P. Kiesmüller, 2022; Rippe and Gudrun P. Kiesmüller, 2023a). Despite these contributions, existing approaches typically rely on structured or codified input data, overlooking the unstructured textual fault descriptions that are often not only available prior to a field technician's intervention, but also represent the sole information describing the issue. The increasing availability of digital maintenance records presents a valuable opportunity to move beyond rule-based heuristics toward predictive, data-driven methodologies. In this context, our previous work introduced CBR-SPP (Schiavo et al., 2024), a Case-Based Reasoning system that leverages historical repair data to recommend spare parts based on textual similarity and appliance metadata. Although effective, CBR-SPP is limited by its reliance on case retrieval and exhibits reduced generalization capabilities when faced with previously unseen appliance-failure scenarios.

In this paper, we propose the DL-SPP (Deep Learning Spare Part Prediction) system, which predicts the parts needed for a repair task using both structured appliance information and semantic representations of free-text fault descriptions. The task is formulated as a multi-label classification problem, and predictions are generated through a transformer-based deep learning architecture. Unlike prior approaches that assume fixed demand probabilities for each part, DL-SPP learns these probabilities directly from data by modeling the relationships between appliance types, fault symptoms, and historically used spare parts.

Our experimental evaluation on a real-world dataset of household appliance repairs confirms the advantages of this approach: DL-SPP significantly outperforms CBR-SPP across all evaluation metrics, achieving a greater reduction in multi-visit repair tasks. These results underscore the value of incorporating unstructured information into spare parts planning, and demonstrate the practicality and scalability of DL-SPP as a real-time decision support tool in field service logistics.

The remainder of the paper is organized as follows. Section 2 formalizes the problem. In Section 3 we describe the proposed DL-SPP system. Section 4 presents the experimental setup and results. In Section 5 we draw some conclusions.

2. PROBLEM STATEMENT

The Spare Part Prediction (SPP) problem addressed in this work focuses on predicting the set of spare parts required by Field Service Technicians (FSTs) to successfully complete household appliance repair interventions. The objective is to anticipate, prior to the FST's visit, the parts most likely needed, thereby preventing incomplete repairs and reducing the number of repeat visits. We assume the availability of a historical archive where each record contains the following standard information typically gathered during fault management.

- *Appliance Type* (*AT*): a categorical alphanumeric code that identifies the type of appliance involved in the intervention. Each code groups together appliances with similar structure and functionality, typically belonging to the same product line or technical platform. For instance, multiple washing machine models that share internal components may be associated with the same *AT* code.
- *Fault Description* (*FD*): an unstructured textual field containing a brief description of the symptoms observed, as communicated by the appliance owner during the initial contact with the service center (e.g., "The door does not close properly" or "Water leaks during the rinse cycle").
- *Spare Parts Used* (*SP*): the set of alphanumeric identifiers of spare parts actually installed by the FST to complete the repair.

This setting is both practical and realistic, as companies providing field repair services commonly acquire such information as part of their routine operations. The recorded data constitute a repository of repair interventions, which can be used to train data-driven models aimed at supporting FSTs by anticipating the spare parts likely required for similar future cases.

To better illustrate the structure of the available data, Table 1: reports a real-world example of a repair intervention record extracted from the historical archive.

Table 1: Example of a repair intervention record.

Appliance Type (AT)	Fault Description (FD)	Spare Parts Used (SP)
76852	The door does not close properly and the program cannot start.	{HK7552, FN3786}

Formally, each input sample is represented as $x = (AT, FD)$, where AT is a categorical variable and FD is a free-text input. The task consists in predicting a set of spare parts $Y_x = \{sp_1, sp_2, \ldots, sp_k\}$, where each $sp_i \in \mathcal{SP}$, the full catalog of n available spare parts.

3. PROPOSED SOLUTION

To effectively address the SPP problem described in the previous section, we propose a deep learning-based SPP system, named DL-SPP. We formulate the task as a multi-label, multi-class classification problem, where each input may be associated with multiple relevant labels (i.e., spare parts). The objective is to accurately predict the set of spare parts required for each intervention, with the aim of reducing multi-visit repair tasks and enhancing overall service efficiency. The system predicts the set of spare parts required for a specific appliance repair task by directly learning relationships between textual fault descriptions, structured appliance metadata, and spare parts usage from historical data.

Our proposed DL-SPP architecture is illustrated schematically in Fig. 1. The model is based on a Multi-Layer Perceptron (MLP) neural network, specifically tailored for multi-label classification, where each spare part is treated as an independent output label. DL-SPP leverages advanced semantic representations derived from Sentence-BERT (SBERT) embeddings for textual fault descriptions. To ensure multilingual compatibility and robustness in handling textual inputs, we adopt the pre-trained *distiluse-base-multilingual-cased-v1* SBERT model (Reimers and Gurevych, 2019).

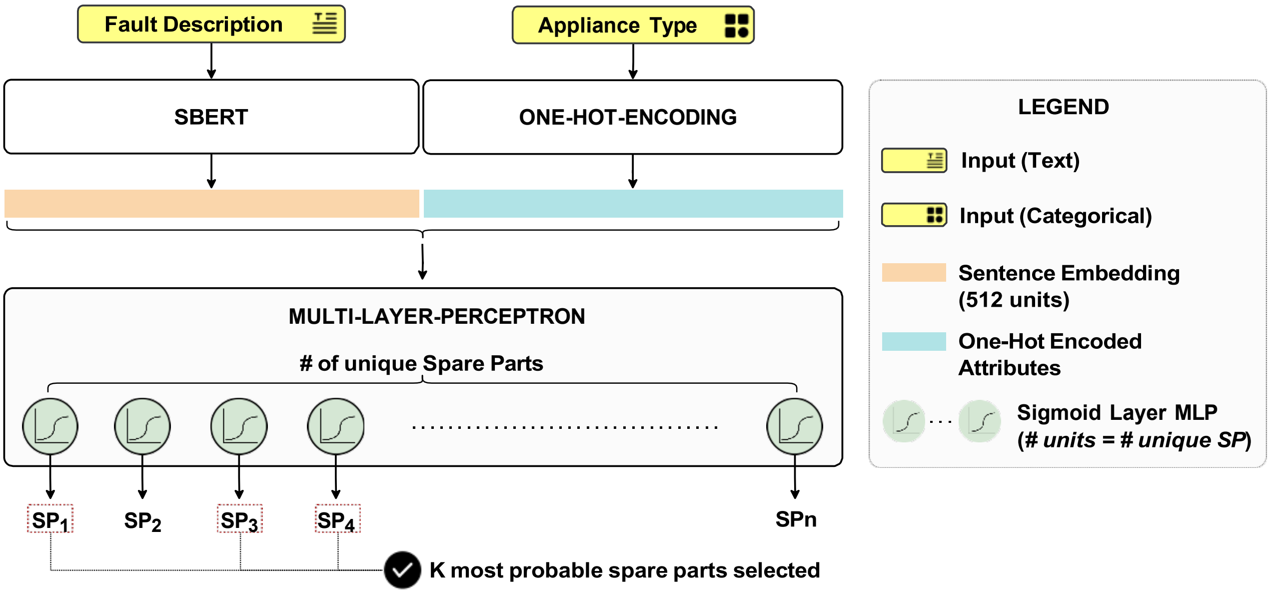

Figure 1: SPP System Architecture Diagram

Given a repair intervention request characterized by its *Appliance Type* and *Fault Description*, we first encode the textual description into a semantic embedding using SBERT. This embedding is a dense numerical representation capturing the semantic meaning of the fault description in a fixed-size 512-dimensional vector. Additionally, the categorical attribute representing the *Appliance Type* is encoded using one-hot encoding, providing structured information about the appliance involved.

The two resulting feature vectors, namely the SBERT embeddings for the *Fault Description* and the one-hot encoded *Appliance Type*, are concatenated into a unified input vector that feeds into the MLP classifier. The MLP consists of multiple hidden layers employing Rectified Linear Unit (ReLU) activation functions to capture nonlinear feature interactions effectively. The output layer uses sigmoid activation functions to predict independent probabilities for each spare part, indicating the likelihood of being required for the specific repair intervention. Formally, the prediction step can be described as follows. Given the input $x = (FD, AT)$, the MLP produces a vector of probabilities $[p_1, \ldots, p_n]$, where p_i represents the predicted likelihood that the spare part sp_i will be required for the repair. The final recommendation list is generated by selecting the top-K spare parts with the highest predicted probabilities, where K is a user-defined parameter balancing prediction coverage and inventory efficiency.

This architectural choice leverages semantic embeddings and structured appliance metadata. It enables the DL-SPP system to generalize effectively by learning intricate relationships between appliance types, fault descriptions, and historical spare parts usage patterns, without relying on explicit case retrieval. By avoiding case-based retrieval, the DL-SPP system ensures scalability and high computational efficiency, making it particularly suitable for dynamic and large-scale operational environments.

In the next section, we describe the experimental setup employed to validate the effectiveness and practical applicability of the DL-SPP system, in comparison with the CBR-SPP baseline.

4. EXPERIMENTAL RESULTS

This section presents the experimental evaluation of the proposed DL-SPP system. We begin by describing the real-world dataset employed in our study, including the preprocessing steps and its key characteristics. Next, we outline the experimental setup and the evaluation metrics used to assess model performance. Finally, we compare DL-SPP with the CBR-SPP baseline through both quantitative analysis and qualitative insights.

4.1. EXPERIMENTAL SETUP

The proposed DL-SPP system was validated using the same real-world dataset introduced in our previous work on the CBR-SPP system (Schiavo et al., 2024). The dataset was provided by a leading household appliance manufacturer and includes approximately 131,000 repair intervention records collected between 2020 and 2023. Each record contains the three key elements described in Section 2: the *Appliance Type*, a categorical code identifying the device class; the *Fault Description*, an unstructured textual description of the failure as reported by the appliance owner; and the set of *Spare Parts* used by the FST to complete the repair.

To ensure data quality, a preprocessing step was applied. Records containing missing or malformed values in either the *Appliance Type* (AT) or *Spare Part* (SP) fields were discarded. *Fault Descriptions* (FD) were retained in their original languages—Italian, German, or French—as the selected SBERT model (`distiluse-base-multilingual-cased-v1`) supports multilingual input. The categorical AT values were encoded using one-hot encoding.

The dataset was split chronologically: repair tasks from 2020 to 2022 formed the training set, while tasks from 2023 constituted the test set. To focus on scenarios where predictive support is most impactful, we excluded single-visit repairs from the test set. These typically involve commonly stocked parts and are less sensitive to prediction quality. Instead, we retained only records of repair tasks that originally required multiple visits.

Table 2 summarizes the characteristics of the dataset, grouping repair tasks by the number of FST visits required to resolve the issue. The bottom row aggregates statistics for tasks requiring two or more visits, which represent the effective test set used in our evaluation.

Table 2: Summary of the dataset used for DL-SPP evaluation. Repair tasks are grouped by the number of FST visits. 4+ and 2+ denote tasks requiring four or more, and two or more visits, respectively. Mean SP and Max SP refer to the average and maximum number of spare parts used.

# Visits	Training Set (2020–2022)			Test Set (2023)		
	# Records	**Mean SP**	**Max SP**	**# Records**	**Mean SP**	**Max SP**
1	72,536 (69%)	1.53	15	17,317 (67%)	1.54	14
2	25,759 (24%)	2.77	15	6,672 (26%)	2.81	15
3	5,813 (6%)	3.32	20	1,573 (6%)	3.37	16
4+	1,099 (1%)	4.88	17	309 (1%)	4.43	14
2+	**32,671 (31%)**	**2.95**	**20**	**8,554 (33%)**	**2.97**	**16**

DL-SPP was trained using the SBERT-generated embeddings for FD and the one-hot encoded representations of AT. The prediction task was formulated as a multi-label classification problem, where the model predicts a set of relevant spare parts for each repair request.

To determine the optimal architecture, we performed hyperparameter tuning using the *Keras Hyperband Tuner* algorithm[1] on a validation subset (20% of the training set). Once the optimal architecture was identified, the model was retrained using the entire training set, including the validation data, to fully leverage the available data. The final configuration consists of two hidden layers with 4,096 and 2,048 units, respectively, each using *ReLU* activation functions. The output layer uses Sigmoid activation to produce independent probability scores for each spare part class. The network was trained to minimize binary cross-entropy loss.

4.2. Evaluation Metrics

We evaluated the DL-SPP system using the same metrics adopted in (Schiavo et al., 2024), which allow comparison with the CBR-SPP baseline and provide insight into both predictive accuracy and operational impact. For varying values of K, indicating the number of recommended spare parts, we report:

- *Closed Task Percentage (CTP_K)*: percentage of tasks successfully completed in a single visit using the top-K predicted parts.

[1] https://keras.io/keras_tuner/api/tuners/hyperband/

- *Saved Visits Percentage (SVP_K)*: percentage of additional FST visits that could have been avoided.
- *Average Recall of Not Closed Tasks ($recallNC_K$)*: average proportion of spare parts correctly predicted by the system in repairing tasks that have not been closed.
- *Average Recall ($recall_K$)*: average proportion over all tasks.
- *Average Precision ($precision_K$)*: average proportion of correctly predicted parts among the top-*K*.

We refer the reader to (Schiavo et al., 2024) for further details.

4.3. Results

We compared the proposed DL-SPP system against the CBR-SPP baseline on the dataset described in the previous section. Table 3 reports the average results across 10 independent training and testing runs. Standard deviations were consistently low (below 0.5) for all metrics and are thus omitted for brevity.

Table 3: Values of the evaluation metrics with respect to increasing values of *K* recommended spare parts, obtained by CBR-SPP and DL-SPP systems. Best model for each *K* is highlighted in bold.

K	Model	CTP_K	SVP_K	$\overline{recallNC_K}$	$\overline{recall_K}$	$\overline{precision_K}$
3	CBR-SPP	15.24	14.32	14.02	30.69	26.41
	DL-SPP	**22.25**	**17.52**	**21.88**	**38.65**	**30.26**
6	CBR-SPP	25.01	23.71	21.62	45.14	20.01
	DL-SPP	**32.97**	**27.56**	**32.57**	**53.09**	**22.27**
9	CBR-SPP	32.57	31.36	26.68	54.35	16.43
	DL-SPP	**40.38**	**36.04**	**37.64**	**61.84**	**17.97**
12	CBR-SPP	38.02	36.75	30.61	60.58	13.96
	DL-SPP	**45.07**	**41.63**	**42.11**	**68.91**	**15.95**
15	CBR-SPP	43.24	41.81	33.52	65.27	12.19
	DL-SPP	**47.71**	**46.82**	**45.05**	**72.37**	**13.96**
18	CBR-SPP	47.46	45.98	35.99	68.86	10.84
	DL-SPP	**53.14**	**51.23**	**47.56**	**76.21**	**12.49**
21	CBR-SPP	51.48	50.12	38.28	71.99	9.78
	DL-SPP	**56.08**	**55.72**	**49.33**	**79.32**	**11.34**

As expected, both systems show performance improvements as *K* increases, confirming their ability to suggest more relevant spare parts when a larger number of predictions is allowed. However, DL-SPP consistently outperforms CBR-SPP across all evaluation metrics and for every value of *K*. Notably, DL-SPP achieves a significant improvement in CTP_K and SVP_K, indicating a greater capability in completing repair tasks with fewer technician visits. These gains are particularly relevant in operational contexts where minimizing repeat visits leads to lower costs and higher customer satisfaction.

For lower values of *K*, the improvements of DL-SPP over the baseline are particularly significant. At $K = 3$, for example, DL-SPP achieves a CTP_3 of 22.25%, compared to 15.24% for CBR-SPP, and reduces technician visits by 17.52%, compared to 14.32%. These results highlight DL-SPP's capacity to produce shorter and more accurate spare part recommendations, especially beneficial for optimizing inventory and service efficiency.

The advantage of DL-SPP remains evident as *K* increases, although the relative gains diminish. For instance, the difference in SVP_K between $K = 18$ and $K = 21$ is about 4 percentage points, indicating that most relevant parts are captured at lower *K* values. This suggests DL-SPP can provide effective recommendations with fewer predicted components, benefiting inventory management and operational efficiency.

The DL-SPP system consistently achieves higher values of $recall_K$ and $precision_K$ across all *K*, confirming improved completeness and accuracy. Precision is especially critical for minimizing overprovisioning, as it reflects the proportion of predicted parts actually required. Although it naturally declines as *K* increases due to less relevant items, DL-SPP maintains consistently higher precision than the baseline, underscoring its effectiveness in generating more targeted and efficient recommendations.

Overall, the results demonstrate that the DL-SPP system provides a robust and scalable solution for spare parts prediction. Its ability to generalize from historical data allows it to effectively learn correlations between fault descriptions and the components required for repair, leading to fewer visits and better resource allocation.

5. CONCLUSIONS

In this paper, we proposed DL-SPP, a novel deep learning-based approach for predicting the spare parts required in household appliance field service operations. By integrating semantic embeddings of fault descriptions with structured appliance type information, DL-SPP addresses key limitations of prior retrieval-based methods, such as the Case-Based Reasoning Spare Parts Prediction (CBR-SPP), particularly in terms of generalization and predictive accuracy. Experimental evaluation conducted on a real-world dataset from a leading appliance manufacturer demonstrated that DL-SPP significantly outperforms the CBR-SPP baseline. Our results indicate substantial improvements in the saved visits percentage, effectively reducing operational costs and enhancing customer satisfaction through more accurate spare part provisioning and minimized technician visits.

The proposed solution exhibits strong scalability and generalization capabilities, essential features for dynamic, large-scale repair environments where spare parts inventories and appliance portfolios evolve continuously. Additionally, by directly modeling relationships between fault descriptions and spare parts usage, DL-SPP facilitates proactive inventory management, optimized technician scheduling, and overall improvement in logistics efficiency, aligning closely with key goals of Operations Research.

Acknowledgments

This work has been partly funded by the PNRR - M4C2 - Investimento 1.3, Partenariato Esteso PE00000013 - "FAIR - Future Artificial Intelligence Research" - Spoke 1 "Human-centered AI" under the NextGeneration EU programme, and by the Italian Ministry of University and Research (MUR) in the framework of the FoReLab and CrossLab projects (Departments of Excellence).

LITERATURE

Bijvank, M., & Vis, I. F. (2010). Lost-sales inventory theory: A review. *European Journal of Operational Research*, *204*(2), 195–205. https://doi.org/https://doi.org/10.1016/j.ejor.2011.02.004

Pham, D. T., & Kiesmüller, G. P. [Gudrun P.]. (2022). Multiperiod integrated spare parts and tour planning for on-site maintenance activities with stochastic repair requests. *Computers & Operations Research*, *148*, 105967. https://doi.org/https://doi.org/10.1016/j.cor.2022.105967

Prak, D., Saccani, N., Syntetos, A., Teunter, R., & Visintin, F. (2017). The Repair Kit Problem with positive replenishment lead times and fixed ordering costs. *European Journal of Operational Research*, *261*(3), 893–902. https://doi.org/https://doi.org/10.1016/j.ejor.2017.02.019

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing*, 3982–3992. https://doi.org/https://doi.org/10.48550/arXiv.1908.10084

Rippe, C., & Kiesmüller, G. P. [Gudrun P.]. (2023a). The repair kit problem with imperfect advance demand information. *European Journal of Operational Research*, *304*(2), 558–576. https://doi.org/https://doi.org/10.1016/j.ejor.2022.04.019

Rippe, C., & Kiesmüller, G. P. [Gudrun P]. (2023b). The added value of advance demand information for the planning of a repair kit. *Central European Journal of Operations Research*, *31*(1), 311–335. https://doi.org/https://doi.org/10.1007/s10100-022-00809-1

Saccani, N., & Johansson, P. (2017). Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *Omega*, *47*, 185–199. https://doi.org/https://doi.org/10.1016/j.omega.2011.06.008

Schiavo, A., Marcelloni, F., Ducange, P., Renda, A., Ruffini, F., & Dani, R. (2024). On Predicting Spare Parts for Field Services by Leveraging Fault Description and Historical Repairing Data. *IEEE Access*, *12*, 162756–162768. https://doi.org/10.1109/ACCESS.2024.3487493

Smith, S. A., Chambers, J. C., & Shlifer, E. (1980). Note—Optimal Inventories Based on Job Completion Rate for Repairs Requiring Multiple Items. *Management Science*, *26*(8), 849–854. https://doi.org/10.1287/mnsc.26.8.849

Teunter, R. H. (2006). The multiple-job repair kit problem. *European Journal of Operational Research*, *175*(2), 1103–1116. https://doi.org/https://doi.org/10.1016/j.ejor.2005.05.023