Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 277 (2026) iii-xxvi

Procedia Computer Science

www.elsevier.com/locate/procedia

Table of Contents

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 277 (2026) 91–100

Procedia Computer Science

www.elsevier.com/locate/procedia

7th International Conference on Industry of the Future and Smart Manufacturing (former International Conference on Industry 4.0 and Smart Manufacturing)

The use of Natural Language Processing (NLP) in aviation: A case study using BERT

Mariana M. Lopes[a,*], Duarte Dinis[a,b,c], Roberto Sala[d], Fabiana Pirola[d]

[a]*Department of Mechanical and Industrial Engineering, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal*

[b]*UNIDEMI, Department of Mechanical and Industrial Engineering, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal*

[c]*CEGIST, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisboa, Portugal*

[d]*Department of Management, Information and Production Engineering, University of Bergamo, Viale Marconi, 5, 24044 – Dalmine (BG) – Italy*

Abstract

In the aircraft Maintenance, Repair and Overhaul (MRO) industry the high uncertainty of resources needed for a given operation is a very recurrent problem. This implies that being able to collect and analyse data from past events to forecast future ones provides a company with more stability of outcomes and a competitive advantage in the industry. However, the heterogeneous manner in which data is collected and stored causes it to be often disregarded or at least not used in all its potential. This paper presents the development and evaluation of a BERT-based Natural Language Processing (NLP) model designed to classify aircraft zones from free-text maintenance reports. Using maintenance reports of 372 maintenance projects from a Portuguese MRO, the model achieved an accuracy of 85.69%. While developing the model a sensitivity analysis was performed to examine the impact of learning rate, warm-up period, and dropout probability on performance, allowing considerations for future model development in this area. In addition, the added value of this paper includes recommendations for maintenance management in companies intending to implement such model.

Keywords: Aircraft Maintenance; Maintenance Management; Natural Language Processing; BERT

1. Introduction

The unpredictable nature of maintenance demand poses significant challenges for capacity planning and resource allocation of Maintenance, Repair and Overhaul (MRO) activities organizations. Companies pursue the ability to use past data to forecast future events. In maintenance, past data is stored in maintenance reports referring to performed tasks. In these reports, several types of fields are found, namely Numeric, Date/Time, Free Text, Boolean,

Dropdown/Select Fields, among others. The data from the numeric/dropdown/Boolean fields can be used for forecasting, but are often found filled incorrectly, or not filled at all, which makes it impossible to use without treatment. Unstructured free-text fields in maintenance reports often contain valuable insights about the tasks performed and, among other things, are the key to correct these mistakes. However, they remain underutilized due to the impracticality of manually processing them. Recent advances in Natural Language Processing (NLP), particularly transformer-based models, offer a promising solution by automating text classification tasks with state-of-the-art accuracy.

In the context of Industry 4.0 (and recently 5.0), these data-driven approaches play a central role in advancing smart manufacturing systems. Specifically, NLP-based classification of maintenance reports can act as a key enabler of predictive maintenance pipelines, where structured insights from historical data are essential to anticipating future failures. By situating this study in the smart manufacturing ecosystem, the contribution extends beyond aviation maintenance to demonstrate how NLP can accelerate the transition of the paradigms of Industry 4.0 and 5.0. This study focuses on the following: 1) the practical use case of zone classification in aviation maintenance; 2) the fine-tuning and hyperparameter optimization of BERT; and 3) the analysis of results and recommendations for industrial implementation.

The paper is organized as follows: Section 2 is a quick overview of related work found; Section 3 explains the use case, dataset and development of the model; Section 4 presents results and sensitivity analysis; and Section 5 presents the conclusion and recommendations.

2. Related Work

2.1. BERT

BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained model that reads text in both directions, which develops a deeper context understanding. Built on Transformer encoder layers, it replaces sequential processing of recurrent neural networks (and other similar models) with parallel self-attention, the mechanism to focus and relate different parts of the same sequence [1]. It has two training phases, pre-training and fine-tuning. When pre-training the model is introduced to large text corpora and learns patterns in language, during it, BERT uses Masked Language Model (MLM), randomly hiding tokens to predict them, and Next Sentence Prediction (NSP) to learn inter-sentence relationships. Whereas in fine-tuning the model is presented with task-specific data further increasing its capacities.

2.2. NLP in Aviation Maintenance

In the maintenance field, the skills required to understand and solve a problem are still regarded as a valuable resource gained through a great deal of experience. Following this principle, and especially in aviation, the information coming from data of previous maintenance work should be considered a valuable asset to tackle the uncertainties in workload prediction, among other problems. Experts are aware of this, and references already exist in the literature [2] [3]. Jing et al. [4] present BERT as a "powerful tool for understanding contextual information" that achieved "impressive results in numerous language understanding tasks" and this is a general sentiment in literature about NLP tools. With such tool, the problem of initial treatment of data can be tackled, incorrect or inexistent numeric/dropdown/Boolean fields can be corrected, and historical data-driven forecasting tools can be developed.

NLP is the specific branch of AI "concerned with processing natural languages such as English or Mandarin. This processing generally involves translating natural language into data (numbers) that a computer can use" Amin et al. [5] present a review about previous research on the application of NLP in aviation. In it, three main focus areas in NLP research for aviation are identified: Aviation maintenance; safety reports analyses; and air traffic control. The paper includes literature from 2010 to 2022 and eleven "NLP Software Products in the Aviation Industry" are catalogued. These software products are described as NLP solutions which automate the process of text analysis, specifically for the aviation industry. The paper concludes that the offers for software in the field is still limited, but they are growing in number. From the eleven software products found, 7 are for maintenance, 2 for safety reports, and 5 for air traffic control. When stating the areas for the tools, an additional area is catalogued, named “customer

interactions", with 4 NLP software tools available. From the 7 found for aviation maintenance, 6 are intended for text classification, the same as the model developed in this work. This is more than half of the software found, which indicates that there is interest in the topic. Other functions found in literature for NLP software are: natural language understanding; natural language generation; topic modelling; information retrieval; semantic similarity assessment; document summarization; speech recognition; and machine translation. For the maintenance field, Amin et al. [5] found 2 applications of NLP: support for switching to predictive maintenance; and assisting MRO technicians in the maintenance procedures themselves. The use of NLP in aviation maintenance helps decrease the time for text analysis, making it more time efficient. This allows for the exploitation of important historical data and helps in future decision making driven by actual past records, which is more reliable and more accurate than simulation, for example.

However, Brundage et al. [2], refers that exploring the potential of NLP to increase maintenance and maintenance management know-how is falling behind compared to other fields. The authors explain that the use of NLP in maintenance is not an extremely innovative action, but, on the contrary, a bit of a copy of other use cases, like the healthcare field, for example, and something to catch up on. This is a general sentiment in literature expressed in [4] or [6] among others. Amin et al. [5] also refer that application of NLP in maintenance is far from its full potential. The development of such tools is delayed because of the complexity of domain-specific vocabulary or abbreviations [2] [5] [7]. This problem is very well presented in [2], which argues that researchers develop NLP pipelines with non-technical language in mind, and there is a lack of adaptation of NLP for a technical and specific environment. Since day-to-day language is too different from the technical one, and the industrial systems currently developed are based on non-technical language processing pipelines, this fact can undermine results and delay implementation.
In some cases, however, it is already possible to see domain experts, analysts and researchers teaming up to create adequate pipelines. One such case is [8], where the effort is in the creation of a collaborative open-source library of technical and domain-specific language datasets, including aviation. Other similar works include Tikayat Ray et al. [9], which developed an annotated aerospace requirements corpus for requirements classification. Another promising effort is the work of Chandra et al. [10] which develops Aviation-BERT, a BERT model with a domain specific pipeline, and that of Sala et al. [11].

3. Use Case and Model Development

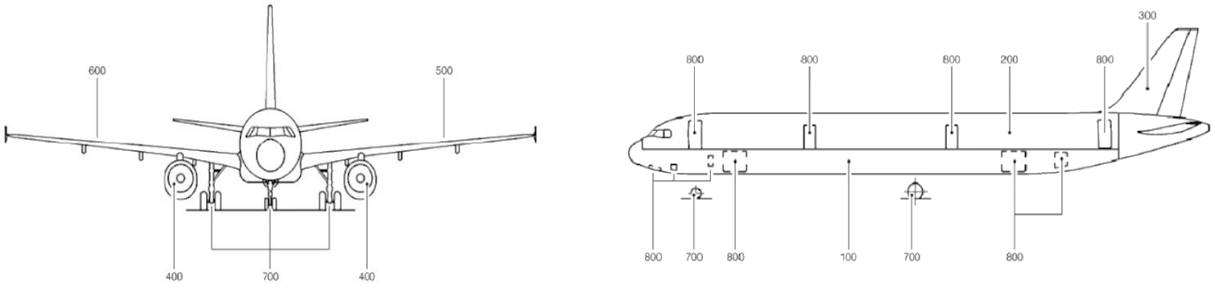
3.1. Use Case

The host MRO organization intends to develop a tool for forecasting based on previous maintenance reports. Data-driven forecasting tools use statistics obtained from clear numerical/dropdown/Boolean data. However, to ensure accurate predictions of future workload, data inputs should be correct, which is not always the case. For this work, the goal was the accurate classification of the aircraft zone to be maintained. This specific field had previously been manually corrected by a field expert and only 6.37% were found to be totally correct. The result of the corrections made by the expert was available in the used dataset, hence the model could learn in a supervised manner. BERT is fine-tuned to the free text fields of said reports according to each correct zone as labels in order to obtain a model capable of classifying aircraft zones. Zone classification is an industry standard, enforced in Europe through the European Union Aviation Safety Agency's (EASA) continuing airworthiness regulations and comprises a three-digit code with each consecutive number being a smaller, more specific area of aircraft [12]. Fig. 1 shows the Major Zones described by the first digit.

3.2. Dataset

Due to field requirements, aircraft maintenance is extensively documented. Every maintenance project is divided in tasks, and each of those tasks is documented in its own report, which means specific and useful information can be retrieved from historical data. The used dataset was received in an Excel file containing a column for each of the report's fields and a row for each maintenance task from each maintenance project. The dataset contained reports for 372 maintenance projects, as they are called in the data, collected between September 2002 and December 2015 in an MRO company [12]. These 372 projects translate into 62508 tasks reported and 62508 rows of the dataset.

Furthermore, thirteen queries/report fields make thirteen columns of data. The dataset had the shape of 62508x13, but it was further treated before being fed into the model.



Major Zone Code	Description
100	Lower fuselage (below the floor line)
200	Upper fuselage (above the floor line)
300	Tail cone and empennage group
400	Powerplant and nacelles
500	Left half-wing
600	Right half-wing
700	Landing gear wheel well
800	Doors (main, service, cargo, emergency)

Fig. 1. Major zone codes identified in the dataset

3.3. Model Development

This step is summarized in the scheme of Fig. 2.

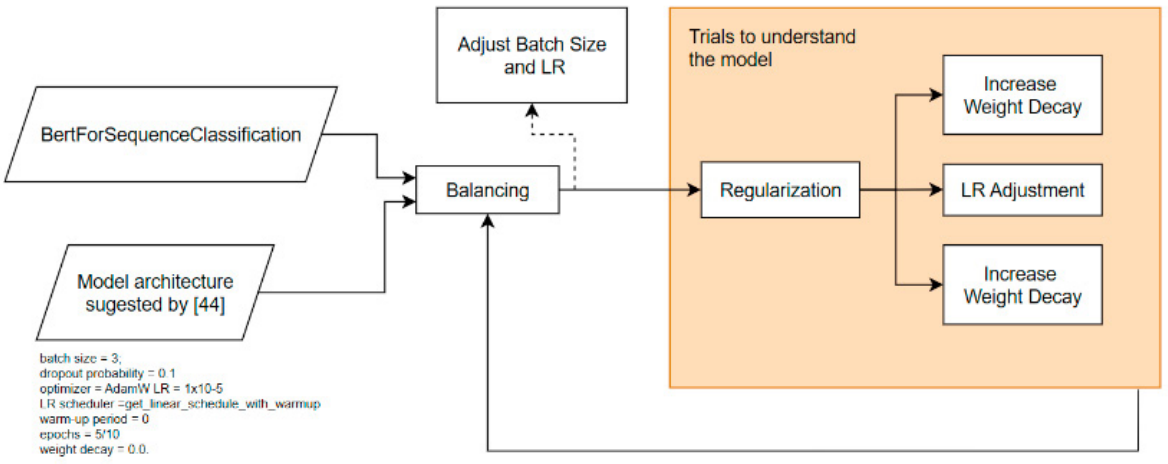


Fig. 2. Model changes in the exploring stage

From the whole dataset only four columns and 11567 lines are used, so the input for the model is a 11567 x 4 matrix. Furthermore, these 11567 free text entries were randomly split into training and validation datasets in a 80-20 ratio. During pre-processing, a decision was made to classify tasks only until the second digit of the zone code, in order to avoid unbalance in number of samples between labels, leaving the third digit to further studies. The chosen

tokenizer and encoder is *BertTokenizer* from the *Transformers* library developed by *HuggingFace*. For model implementation, an initial good set of hyperparameters was found to be:

- Batch size = 16
- LR = 1x10-5
- LR curve: Warm-up of 10% and a decay phase at the end
- Weight Decay = 0.01
- Dropout probability = 0.1
- Optimizer = AdamW
- LR scheduler

4. Results and Sensitivity Analysis

The sensitivity analysis was performed according to Fig. 3.

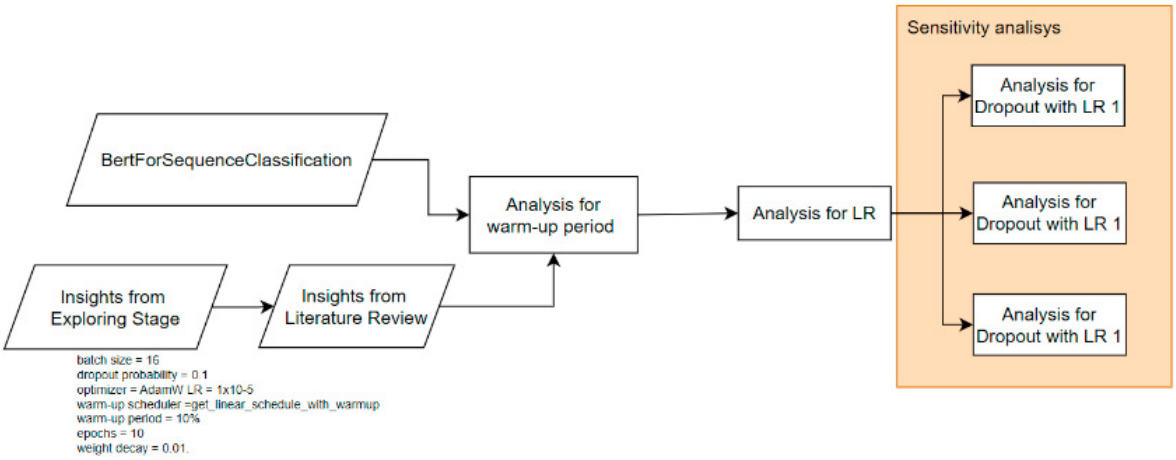
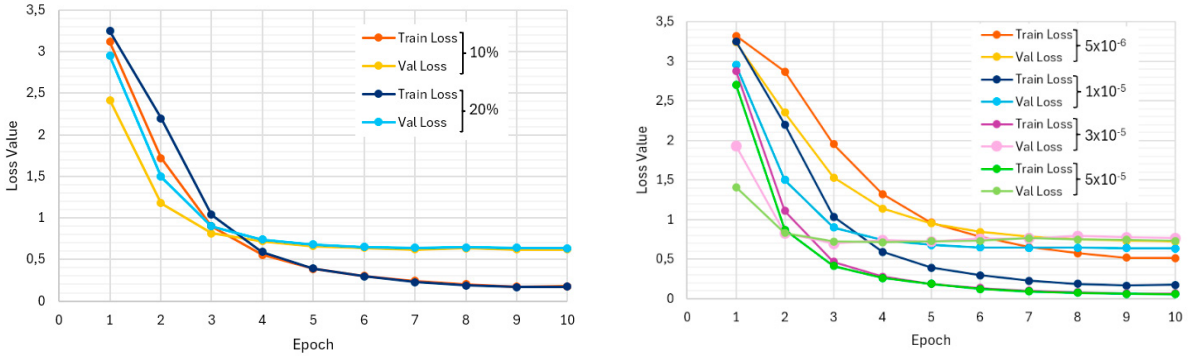


Fig. 3. Model changes in the hyper-parameter search stage

Fig. 4 and Fig. 5 represent the curves of progression of loss throughout epochs when training the model and validating results. They are useful to understand how the model reacted to the various levels chosen for each parameter. In each graph, same colour curves represent same levels of parameter and dark shades represent training loss, whereas lighter shades the validation loss. This facilitates comparisons between the same metrics and between different parameter setups.

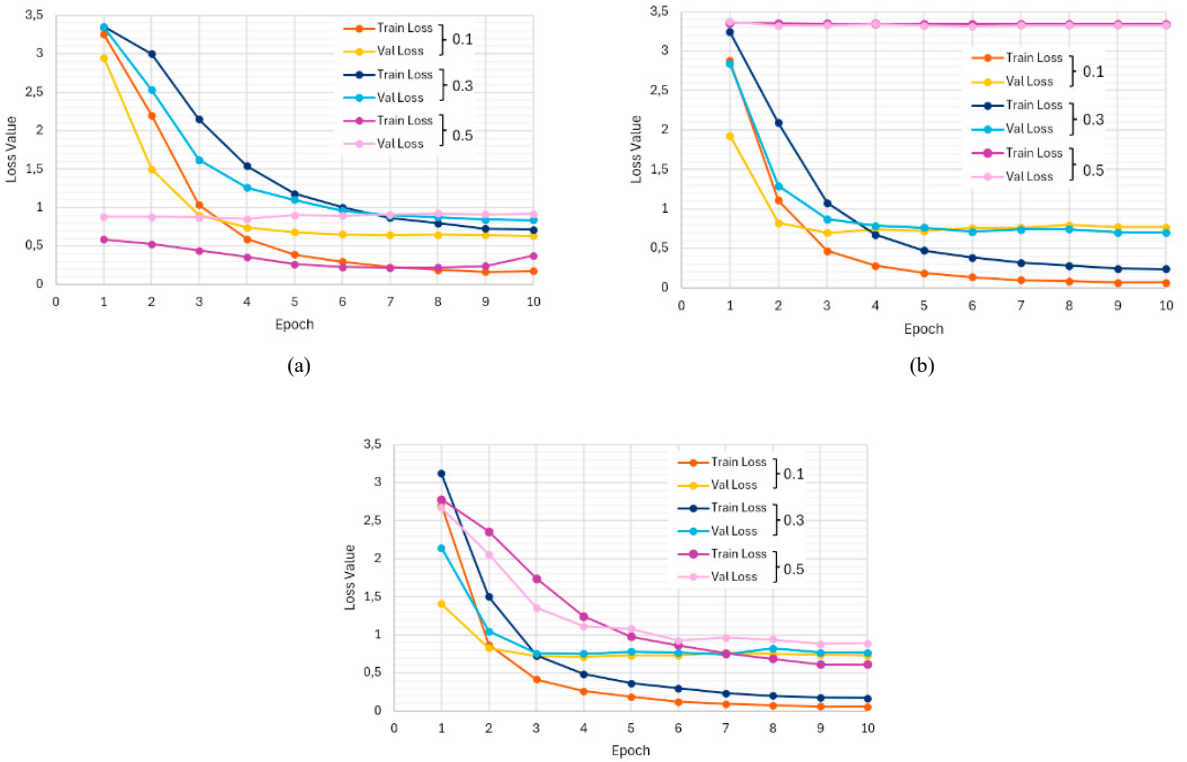
The initial setup of parameters was the one found in Section 3.3. After those initial values, some parameters and their levels were chosen to the study based on previous works, with BERT, both in and out of the field of aviation maintenance, found in a literature review. The first parameter chosen was the Learning Rate (LR) warm-up period. As one can see in Fig. 4 (a), both warm-up periods showed similar progression throughout epochs, which implies that in this range, model and number of epochs, the length of the warm-up period is not relevant. Nevertheless, the accuracy in the best epoch was slightly higher for the 20% warm-up period, with 84.36% and 85.69%, for 10% and 20%, respectively. Second, a study on different LR was conducted to understand what range would be of interest. In Fig. 4 (b) it is noticeable how different LR values affect the model much more than the warm-up periods. Not only do they start and end at different levels, but also their patterns are different. For the lower LR, losses start at higher values and never reach the same as other levels of LR, making it not ideal. LRs $3x10^{-5}$ and $5x10^{-5}$ reached a very low training

loss, but the validation loss curve had its minimum early, in third and fourth epochs, respectively. This can be observed in Figure 4b, which can indicate their adequateness since they reached a peak within or training time. Lastly, LR of $1x10^{-5}$ reached low levels on both curves and did not have any instability, i.e., both curves decreased normally. The accuracies for those setups were 82.54%, 85.69%, 82.80% and 84.40%, in ascending order of LR.



(a) (b)

Fig. 4. (a) Training and Validation losses with different Learning Rate (LR) warm-up periods; (b) Training and Validation losses with different LRs.



(a) (b)

(c)

Fig. 5. Training and Validation losses with different dropout levels and a LR of: (a) $1x10^{-5}$; (b) $3x10^{-5}$; and (c) $5x10^{-5}$

Next, a sensitivity analysis was performed with three values for both dropout and LR (the lower LR of $5x10^{-6}$ was disregarded). Loss curve graphs can be seen in Fig. 5 and accuracy for the best epoch in Table 1. In Fig. 5 (a), the LR is the same as the previous configuration, $1x10^{-5}$, which included the dropout at 0.1. There were no signs of over-fitting, so it was not expected that the higher dropout levels would yield a better performance and that was, in fact, the case. In fact, for all levels, the model behaved mostly as expected: with the two lower LR, Fig. 5 (a) and Fig. 5 (b), the higher dropouts cause instability since low LR does not call for much compensation in "forgetting" some nodes. The highest LR, Fig. 5 (c), has a more aggressive learning pattern, being prone to wrong extrapolations. This means that a higher dropout causes less severe instability and higher accuracy.

Table 1. Accuracy for all levels of sensitivity analysis with three different Learning Rate values.

		Learning Rate		
		$1x10^{-5}$	**$3x10^{-5}$**	**$5x10^{-5}$**
Dropout probability	**0.1**	85.69	82.62	84.40
	0.3	78.13	85.09	84.52
	0.5	80.25	11.66	77.09

The most notorious conclusion is that the 0.5 dropout, the highest, is not favorable in any level of LR. The best performances were the combinations: low LR plus low dropout; medium and high LR plus medium dropout. Both the medium and high fitting well with the same dropout may be associated with the fact that $5x10^{-5}$ LR is still inside what is commonly considered a "normal" range for the LR and may still not be high enough for a high level.

´For the best performing model the best setup is then a 20% warm-up rate paired with $1x10^{-5}$ LR and 0.1 dropout. With an accuracy of 85.69% and an F1-score of 85.79% the confusion matrix for the validation dataset is in Fig. 6. From it labels with low individual accuracy and pairs of labels with high absolute number of wrong predictions were discovered.

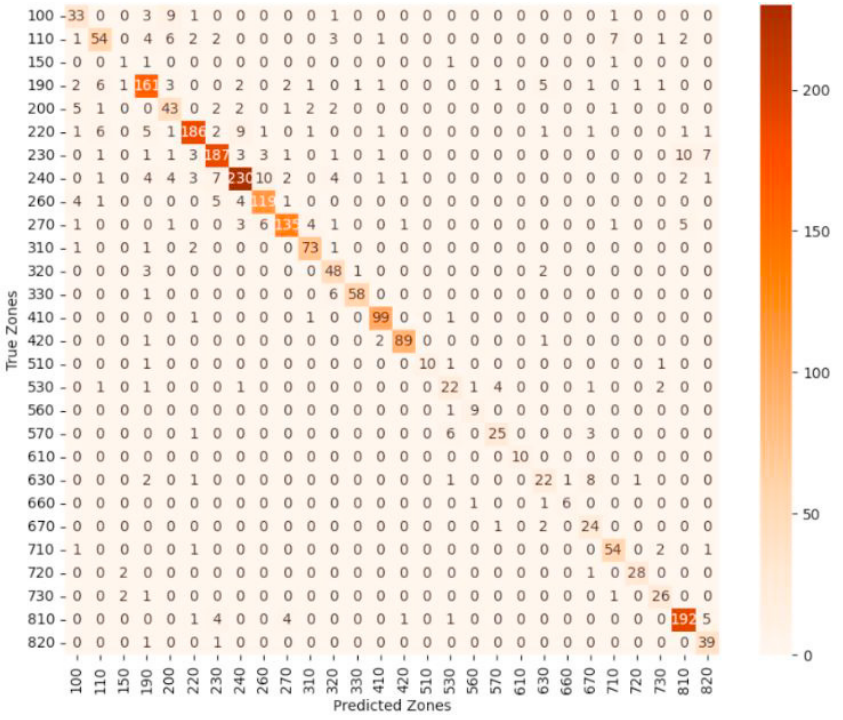


Fig. 6. Confusion Matrix of chosen model configuration with predicted zones vs true zones

5. Discussion and Recommendations

5.1. Low Accuracy Zones

The computation of accuracy values per class revealed that several zones had considerably lower performance. In particular, zones under the major fuselage labels—100 (Lower Fuselage), 110 (Fuselage Forward Section I), 150 (Fuselage Centre Section III), and 200 (Upper Fuselage)—displayed significant confusion. This can be attributed to the fact that they are all subparts of the fuselage, meaning text descriptions will use similar words and they all share structural similarities, so the model struggles to identify unique features. Also, since these zones describe both interior and exterior areas, the model could struggle to find patterns.

The issue is further complicated by the presence of only two major zones, 100 and 200, in the training data. These zones represent aggregate categories, while their subzones (e.g., 110–190 for 100) offer more detail. The lack of distinct boundaries in description between major zones and their subzones introduces noise into the classification task.

Additionally, zones 530 and 630, which refer to the left and right wings, respectively, comprehensibly share identical descriptions, further confusing the model. Other zones with a relatively low-class accuracy included subzones 570, 660 and 510 which could have the same problem, although since their counterparts did not show the same pattern, it is not possible to generalize the conclusion.

5.2. Pairs with High Number of Errors

Beyond individual zone accuracy, certain zone pairs consistently yielded a high number of absolute misclassifications. For example, the model frequently confused zone 230 (Fuselage Center Section I) with 810 (Fuselage LH side - Doors) and 820 (RH side - Doors). These errors are attributable to overlapping descriptions, such as the mention of doors, seals, and fuselage components that appear in both zones.

Similarly, zones 100 and 200 (lower and upper fuselage) were often confused due to similar vocabulary describing general fuselage issues, such as corrosion, paint, or antenna installation. Pairs like 630 and 670 (Wing interior and trailing edge), and 570 and 530 (left wing trailing edge and interior), also presented confusion stemming from their close physical proximity and interchangeable terms like "spoiler" and "fairing."

Other recurrent confusions include:

- 110 and 710: Similar terminology due to overlap between fuselage and nose landing gear compartments.
- 240 and 230: Similar descriptions of central fuselage sections.
- 330 and 320: Difficulty in differentiating horizontal vs vertical stabilizers due to similar design and problem descriptions.

5.3. Recommendations

A. Rearrangement or Creation of New Labels for Fuselage

Given the high degree of lexical similarity and overlapping tasks in fuselage-related zones, one recommendation is the strategic regrouping of these labels. If regulatory and operational constraints do not allow a definitive rearrangement, then reconfiguring these labels into functionally clustered groups could improve model clarity. Alternatively, if MRO have the need to distinguish current labels, fundamental differences should be identified, and technicians explicitly instructed to differentiate descriptions between adjacent fuselage zones to improve input clarity.

B. Possibly Not Including Major Zones

To reduce label ambiguity, the exclusion of major zones 100 and 200 from the training dataset is suggested. These zones aggregate subzones without offering distinct descriptions, thus confusing the classifier. It would probably be beneficial to exclude these. If the maintenance task refers to the full major zone, then this solution would have to include multi-label classification in order to allow for all the subzones to be selected as labels, this model does not include multi-label classification due to computational constraints.

C. Clarification of Zoning for Wings

For wing zones, the key issue lies in lexical and structural overlap between adjacent or symmetric zones, and probably not from confusion between left and right. Labels such as 530/630 and 570/670 should either be differentiated more clearly in documentation or regrouped where distinctions are functionally irrelevant, just like for fuselage zones. Furthermore, encouraging consistent terminology in maintenance descriptions, especially specifying side (e.g., LH/RH, leading/trailing), can also mitigate misclassifications.

D. Instruction for More Standardized Text

Naturally, when filling the free text fields, a vocabulary that is not only more standardized but also in accordance to the existing zoning, when describing tasks, would help the model in classification. This is true not only for classification of zones, but also other classification models that could be developed according to MRO needs.

E. Technician Instruction on AI

Since most effective solutions include some degree of behavior change from technicians, it is advisable to educate maintenance personnel on the functioning of AI models and the value of accurate, detailed descriptions. This can prevent resistance to procedural changes and foster better alignment between human and machine-readable data.

F. Developing a More TLP-Friendly Environment

Adopting Technical Language Processing (TLP) principles (2) is essential for improving NLP model performance in aviation contexts. Therefore, all companies should be encouraged to, first, contribute to widely available data banks specific for the development of such tools. Second, on a company level, the creation of a dictionary of terms and/or a customized tokenizer, including a whole adapted pipeline like the one in Aviation-BERT (10) is advised.

G. Double Confirmation

While the model achieved an accuracy of 85.69%, the remaining proportion of misclassifications could potentially carry implications in the highly demanding reality of the sector, this is something to be checked in further investigations.

However, if needed, given the model accuracy and depending on acceptable thresholds for this context, it could be deployed in operational systems paired with supplementary human oversight. A dual-check system where model predictions are cross-verified with technician entries, and mismatches flagged, can provide a practical balance between automation and reliability. This ensures high precision without compromising trust in AI predictions.

6. Conclusion

This study applied a fine-tuned BERT model to classify maintenance reports by aircraft zone, addressing persistent inaccuracies in maintenance reports. With an achieved accuracy of 85.69%, far surpassing the 6.37% baseline from manual entries, the model demonstrates strong potential for automating zoning tasks in aviation MRO operations.

For model development, the 20% LR warm-up period, $1x10^{-5}$ LR and 0.1 dropout were found to be the most adequate levels for finetuning this dataset. Thus, the lower level of LR paired with the lower level of dropout probability proved to perform best, followed by the medium level of dropout with LR levels at medium and high (presumed to still be in the range of medium, i.e., not high enough).

Maintenance management wise, the model faced challenges in distinguishing zones with overlapping language, particularly among fuselage and wing labels. These findings support the need for domain-adapted pipelines such as models like Aviation-BERT, as well as more structured technician input.

Key recommendations include revising zone labels, excluding ambiguous major zones, and in the future adopting multi-label classification, and introducing human-in-the-loop validation for critical outputs. This approach not only improves reliability but ensures scalability within regulatory constraints.

Finally, these results positions NLP not just as a tool for efficiency but as a fundamental component of smart manufacturing pipelines. The ability to convert free-text maintenance report fields into structured machine-friendly inputs opens the way towards digital integration, a goal of the industry 4.0 and 5.0.

Acknowledgements

Duarte Dinis acknowledges the Portuguese Fundação para a Ciência e a Tecnologia (FCT) for its financial support via the project UIDB/00667/2025 and UIDP/00667/2025 (UNIDEMI). Duarte Dinis also acknowledges FCT for the financial support under the project UIDB/00097/2025 and UIDP/00097/2025 (CEGIST).

References

[1] A. Vaswani et al., “Attention Is All You Need,” Aug. 2023.

[2] M. P. Brundage, T. Sexton, M. Hodkiewicz, A. Dima, and S. Lukens, “Technical language processing: Unlocking maintenance knowledge,” *Manuf Lett*, vol. 27, pp. 42–46, Jan. 2021, doi: 10.1016/j.mfglet.2020.11.001.

[3] D. Berkholz, M. Schmidt, and P. Nyhuis, “From Volatile Maintenance Data Forecasting to Reliable Capacity Planning ,” in *Proceedings of the World Congress on Engineering and Computer Science 2008*, WCECS 2008, October 22 - 24, 2008, San Francisco, USA, 2008.

[4] X. Jing, A. Chennakesavan, C. Chandra, M. V. Bendarkar, M. Kirby, and D. N. Mavris, “BERT for Aviation Text Classification,” in *AIAA AVIATION 2023 Forum*, Reston, Virginia: American Institute of Aeronautics and Astronautics, Jun. 2023. doi: 10.2514/6.2023-3438.

[5] N. Amin, T. L. Yother, M. E. Johnson, and J. Rayz, “Exploration of Natural Language Processing (NLP) Applications in Aviation,” *Collegiate Aviation Review International*, vol. 40, no. 1, 2022, doi: 10.22488/okstate.22.100211.

[6] E. Hastings, T. Sexton, M. P. Brundage, and M. Hodkiewicz, “Agreement Behavior of Isolated Annotators for Maintenance Work-Order Data Mining,” *Annual Conference of the PHM Society*, vol. 11, no. 1, Sep. 2019, doi: 10.36001/phmconf.2019.v11i1.791.

[7] S. M. R. Naqvi, M. Ghufran, C. Varnier, J.-M. Nicod, K. Javed, and N. Zerhouni, “Unlocking maintenance insights in industrial text through semantic search,” *Comput Ind*, vol. 157–158, p. 104083, May 2024, doi: 10.1016/j.compind.2024.104083.

[8] F. Akhbardeh, T. Desell, and M. Zampieri, “MaintNet: A Collaborative Open-Source Library for Predictive Maintenance Language Resources,” May 2020.

[9] A. Tikayat Ray, B. F. Cole, O. J. Pinon Fischer, R. T. White, and D. N. Mavris, “aeroBERT-Classifier: Classification of Aerospace Requirements Using BERT,” *Aerospace*, vol. 10, no. 3, p. 279, Mar. 2023, doi: 10.3390/aerospace10030279.

[10] C. Chandra et al., “Aviation-BERT: A Preliminary Aviation-Specific Natural Language Model,” in *AIAA AVIATION 2023 Forum*, Reston, Virginia: American Institute of Aeronautics and Astronautics, Jun. 2023. doi: 10.2514/6.2023-3436.

[11] R. Sala, F. Pirola, G. Pezzotta, and S. Cavalieri, “NLP-based insights discovery for industrial asset and service improvement: an analysis of maintenance reports,” *IFAC-PapersOnLine*, vol. 55, no. 2, pp. 522–527, 2022, doi: 10.1016/j.ifacol.2022.04.247.

[12] D. Dinis, A. Barbosa-Póvoa, and Â. P. Teixeira, “A supporting framework for maintenance capacity planning and scheduling: Development and application in the aircraft MRO industry,” *Int J Prod Econ*, vol. 218, no. February, pp. 1–15, Dec. 2019, doi: 10.1016/j.ijpe.2019.04.029.