Data Engineering and Applications

Research Paper

Bidirectional GRU-Attention Network for Intelligent Industrial Fault Report Classification

Ghanim Al Kuwaiti[1], Fahad Al Dosari[1] and Khalid Al Shammari[1, *]

[1] Faculty of Engineering, Department of Computer Science, Kuwait College of Science and Technology, 22054, Doha, Kuwait
*Corresponding author: khalid.s@kcst.edu.kw

Abstract. The industry's stated issues are growing more complicated as intelligent manufacturing advances quickly, necessitating more accurate automatic analysis. The aforementioned items must be arranged in order to support data-driven production management, minimise operating disruptions, and enable prompt maintenance. In order to learn thoroughly from a variety of narratives and technical jargon in industrial reports, this work has built a new classification system and incorporated an attention mechanism with a Bidirectional Gated Recurrent Unit (BiGRU). In order to extract features and reduce informational noise, a bespoke pre-processing pipeline is used that incorporates domain-specific tokenisation and embeddings. The suggested approach is thoroughly tested using a large-scale industrial dataset of 18,734 annotated defect reports, which shows notable differences in text length and class imbalance. According to the experimental results, the BiGRU-Attention model outperforms both the neural baseline and the conventional technique in terms of accuracy and macro F1-score. It is also highly effective at identifying uncommon fault kinds. The framework has been succinctly documented in error analysis and is very adept at managing ambiguous expressions. The model is practical and successful as an intelligent fault-detection tool for enhancing industrial maintenance decision-support capabilities, according to the aforementioned findings.

Keywords: *Fault Diagnosis, Deep Learning, Industrial Automation, Text Mining*

Received on 23 October 2025, Accepted on 29 March 2026, Published on 04 April 2026

Introduction

The rapid development of intelligent manufacturing is transforming the conventional industrial model through the widespread use of automation, the industrial Internet of Things (IIoT), and data-driven decision-making [1]. Appropriate equipment maintenance and failure management are necessary to increase economies of scale and keep up with the world's economy's rapid growth [2]. By offering basic information on equipment failures and repair techniques, fault reports help to address the aforementioned issues earlier and increase the efficiency of maintenance operations [3]. Automate the classification of defect reports to assist the manufacturing organization in managing operational risks, reducing unplanned production disruptions, improving product quality, and quickly resolving maintenance issues [4].

Automated classification of manufacturing problem reports continues to face challenges, despite recent advances in machine learning and natural language processing (NLP) [5]. Decision-Making and Support Vector Machines The complexity and diversity of industrial language cannot be fully addressed by trees, which are conventional machine learning algorithms that require manual feature extraction [6]. The difficulties of non-standardized language and context-dependent narrative structures in industrial contexts are beyond the capabilities of early deep learning models, such as CNNs and RNNs, which have shown some success in general text analysis tasks [7]. Automatic comprehension of these reports is hampered by specialised language, inconsistent textual syntax, varying information durations, and noise introduced by human operators [8]. Many classifiers are prone to overfitting, show poor generalisation, and are not very resilient in the face of heterogeneous real-world data since fault data are typically unbalanced and labelled examples are hard to come by [9]. Poor interpretability and a lack of long-distance semantic model’s plague many of the current algorithms for smart manufacturing systems [10].

In light of the aforementioned issues, this research presents a novel method for classifying manufacturing fault reports that combines an attention mechanism with a Bidirectional Gated Recurrent Unit (BiGRU) neural network. By analysing the text sequence in both directions, the novel architecture can learn both local and global semantic relationships while simultaneously concentrating computational resources on important tokens that are more likely to result in failure. In order to enhance feature representation and model interpretability, a model that can more accurately capture the language and structural aspects of industrial fault reports will be developed. Experiments using a variety of public and commercial manufacturing datasets have demonstrated that the BiGRU-attention model outperforms current baseline methods in terms of accuracy, resilience, and interpretability. The remainder of this work is structured as follows: In Section 2, the problem is outlined and related work is introduced; in Section 3, the BiGRU-attention model's architecture is presented in detail; in Section 4, the experimental setup and thorough results are analysed; and in Section 5, a summary of the study and recommendations for further research are provided.

Related Work and Problem Definition

Fault Report Classification Methods

Intelligent manufacturing systems increasingly use the division of fault reports to facilitate comprehensive maintenance forecast and decision-making [11]. In an industrial setting, fault data can be delivered either unstructuredly as a narrative by engineers or operators, or structuredly as a log file with predetermined fields [12]. While automated parsing of structured data is more convenient, many crucial pieces of information, such context and the rationale behind a diagnosis and corrective actions, are not available in this format and must be extracted from unstructured free-text reports [13]. Traditional information retrieval and data analysis techniques typically do not address the particular language, inconsistent idioms, and usage of technical jargon in such sources [14].

In the early days of industrial text analysis, Support Vector Machines and Random Forests were popular categorisation techniques because they were very easy to install and operate [15]. The majority of them are manual techniques that produce model features like term frequency-inverse document frequency (TF-IDF) and keyword statistics [16]. However, the aforementioned approaches are unable to adequately handle the intricate semantics, vague language, and context-sensitive vocabulary of actual industrial fault reports [17]. Furthermore, because of their relative simplicity, these models are unable to account for inter-sentence linkages or long-distance dependencies, which are essential for comprehending complicated failure modes. The limitations of traditional machine learning, such as its limited capacity to generalise to new expressions and lack of adaptability, have become increasingly apparent as the complexity and volume of manufacturing data have grown [18].

Sequence Models and Attention Mechanisms in Text Mining

Due to the limitations of manually developed feature-based techniques, deep learning sequence models have been developed to automatically learn the temporal and sequential links among words in text [19]. Recurrent Neural Networks (RNNs) were an improvement, even if they could describe the context of arbitrarily long sequences. However, gated units have outperformed the original RNN type due to the vanishing gradient issue [20]. By using internal gating mechanisms to preserve crucial information over an extended duration, both Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTMs) networks have accomplished good semantic learning of industrial reports [21]. Additionally, Bi-GRU models have performed well when extracting information from a sequence in both directions. By merging previous and subsequent contexts, they may also more precisely handle expression ambiguity and discover the complete diagnostic context of a multi-stage manufacturing process [22].

Despite the development of these complex gated structures, they are still unable to differentiate between the different levels of significance for error diagnosis among words and phrases, and as a result, they pay uniform attention to all input tokens. Attention processes have somewhat altered the original paradigm and are now employed to balance the varied input data based on the significance of various components in a task [23].

Improved interpretability and classification accuracy are necessary since the industrial sector is likely to demand domain-specific keywords and phrases [24]. By adding attention mechanisms, BiGRU and other models have successfully improved the accuracy of classification tasks for industrial and technical texts, and these models are more suited for managing multi-grained semantic links in actual defect data [25].

Research Gaps and Problem Statement

Automatic classification of manufacturing problem reports still has several flaws, despite modest advancements. The emergence of textual noise, inconsistent labels, and the erratic and intermittent nature of some defect types continue to hinder stable model deployment and training. Because real-world industrial writings are frequently sparse, contain ambiguous descriptions, and have shifting domain nomenclature, both classical and deep learning models still struggle to generalise to and comprehend these texts. In light of the aforementioned issues, this study advances the creation of intelligent maintenance systems by putting forth a novel BiGRU-Attention framework that satisfies the semantic, structural, and practical requirements of manufacturing fault report classification.

Proposed BiGRU-Attention Model

Data Processing Workflow

To guarantee the accuracy and general application of the industrial fault report categorisation system, a high-precision, general-purpose data-processing pipeline must be set up. The data collection for this investigation began with a combination of enterprise-level maintenance archives and real-time equipment defect records, covering a broad range of manufacturing equipment, operating conditions, and production lines [26]. This data comes from a variety of sources that frequently utilise operator-specific technical jargon, lack thorough coverage, have inconsistent text formats, and are riddled with mistakes like typos and superfluous embedded metadata.

General cleaning and noise reduction will be the focus of the first round. Eliminate unnecessary symbols, excess letters, and extra markup. Standardise the format of numerical expressions, time stamps, and units. At this point, the raw corpus was normalised to increase the results' repeatability and lower systematic mistakes brought on by variations in operator input [27].

The production of the narrative was then tokenised using a domain-optimized segmentation technique. As demonstrated above, technical abbreviations, system-specific IDs, and compound equipment codes were handled as indivisible units rather than broken tokens. The text was regularly changed to lowercase to further reduce its size, and context-sensitive punctuation and control symbols that had no semantic significance in the diagnosis were eliminated.

Due to the industrial text's increased complexity and specialisation, an enlarged stop-word filter is required. Rare terms with high semantic significance to specific machine components and issues were not omitted, but frequently recurring but uninformative words were eliminated using a specifically created industrial stop-word list [28]. The key components of the variant technical statements would have been lost in the absence of context-aware lemmatisation and suitable stemming. For uncommon phrases or mistake codes that were statistically significant in diagnostic analysis, some additional processing has been used to minimise information loss.

The text must first be prepared for the neural network in order to employ sequence models. The cleaned and segmented reports were mapped into a vector space using pre-trained Word2Vec embeddings in order to gain a deeper understanding of the unique linguistic characteristics of the manufacturing industry. The combined corpus was then used to optimise the aforementioned models [29]. Parallel experiments using contextualised embedding techniques like BERT have demonstrated good results for lengthier or more ambiguous report narratives; consequently, stable semantic information has been gained at the model level.

Figure 1 illustrates the fundamental evidence for the stability and flexibility of the suggested BiGRU-attention paradigm. By matching high-quality, context-rich embeddings with strict noise management to preserve semantics, all pipeline stages have been demonstrated to reduce model overfitting and enhance classification performance in the presence of irregular, sparse, or anomalous industrial reports [30].

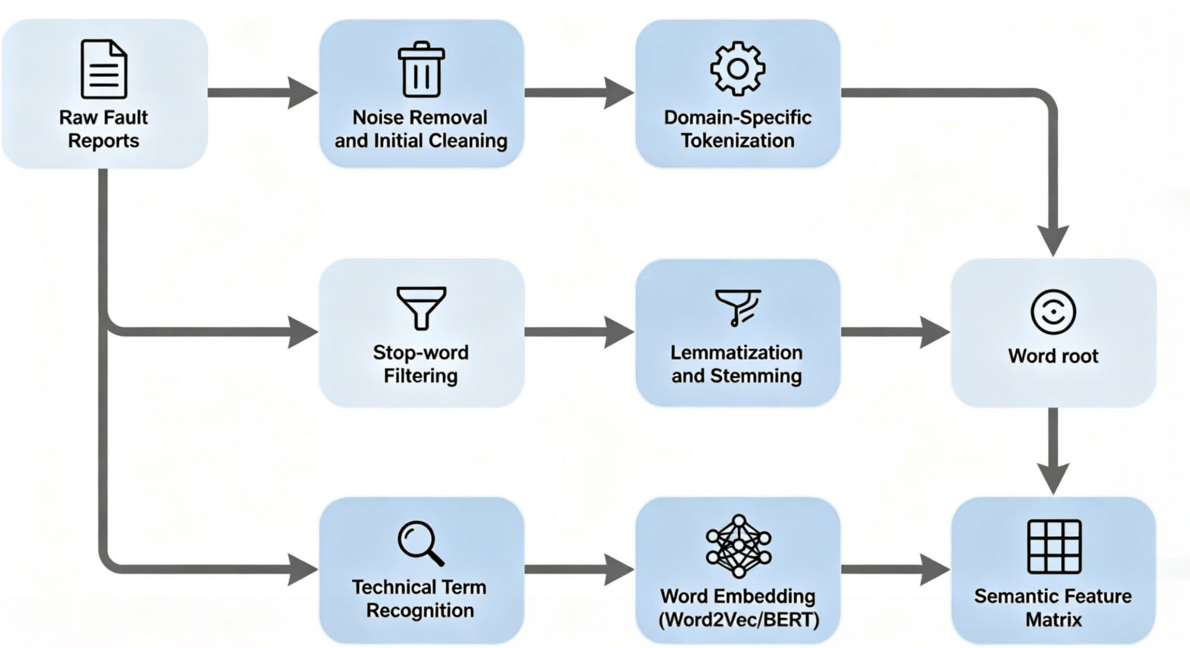


Figure 1. Schematic overview of the manufacturing fault text data processing workflow.

Model Architecture and Core Components

The BiGRU-Attention model developed in this study is expressly tailored to the linguistic and contextual challenges that typify fault reports in complex manufacturing systems.

Following preprocessing, each fault report is transformed into a token sequence $(w_1, w_2, \dots, w_T)$ with length T. Tokens are embedded into dense vectors via pretrained domain-specialized Word2Vec or BERT, assuring coverage of industrial technical terms and synonyms typically missed by general-purpose models.

These embeddings form the input to a bidirectional gated recurrent unit (BiGRU) backbone, designed to capture both past and future context for every token in the sequence. At each timestep t, the token representation passes through forward and backward GRU cells, producing hidden states $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ respectively. The concatenated representation $h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$ preserves comprehensive contextual dependencies across the document. The internal mechanics of the GRU-comprising update (z_t) and reset (r_t) gates-enable flexible memory retention and selective forgetting, updating the hidden state as:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \qquad \text{Eq. (1)}$$

where weights W_*, U_*, and biases b_* are learned during training, σ is the sigmoid activation, and $\odot$ denotes element-wise multiplication.

while the reset gate is given by:

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \qquad \text{Eq. (2)}$$

Next, the candidate hidden state is computed as:

$$\tilde{h}_t = \tanh\left(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h\right) \qquad \text{Eq. (3)}$$

Finally, the new hidden state is updated by:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \qquad \text{Eq. (4)}$$

While BiGRU effectively models sequence-level dependencies, it does not inherently distinguish the varying diagnostic importance of tokens for fault identification. To address this, an attention mechanism is incorporated, enabling the model to learn and amplify the influence of words crucial to specific failure classes. The attention mechanism computes a score for each token:

The attention mechanism computes a relevance score for each token as:

$$e_t = v^{\mathsf{T}} \tanh\left(W_a h_t + b_a\right) \qquad \text{Eq. (5)}$$

These relevance scores are then normalized via a softmax operation to obtain attention weights:

$$\alpha_t = \frac{\exp\left(e_t\right)}{\sum_{k=1}^{T} \exp\left(e_k\right)} \qquad \text{Eq. (6)}$$

The semantic representation of the whole fault report is aggregated as a weighted sum of hidden states:

$$s = \sum_{t=1}^{T} \alpha_t h_t \qquad \text{Eq. (7)}$$

Here, h_t denotes the BiGRU output at position t, W_a, b_a are learned parameters, v is a context vector, α_t is the normalized attention weight, and s is the aggregated semantic representation of the whole fault report, composed by focusing on the most relevant contextual elements. This design allows the model to dynamically prioritize domain-specific error symbols, root cause keywords, or technical terms that might otherwise be lost in a lengthy narrative.

The context vector s is then passed through a fully connected nonlinear transformation:

$$o = \tanh\left(W_o s + b_o\right) \qquad \text{Eq. (8)}$$

$$\hat{y} = \mathrm{softmax}\left(W_y o + b_y\right) \qquad \text{Eq. (9)}$$

where $\hat{y}$ yields the probability distribution over the possible fault categories.

The classifier is trained via categorical cross-entropy loss:

$$\mathcal{L} = -\sum_{c=1}^{C} y_c \log\left(\hat{y}_c\right) \qquad \text{Eq. (10)}$$

where y_c is the true label and C is the number of classes. Regularization (e.g., L_2-norm penalties) is simultaneously applied to mitigate overfitting in realistic, data-sparse industrial scenarios.

The full pipeline-depicted in Figure 2-illustrates how input embeddings, sequential encoding, attention-based aggregation, and class inference are organically chained for end-to-end learning.

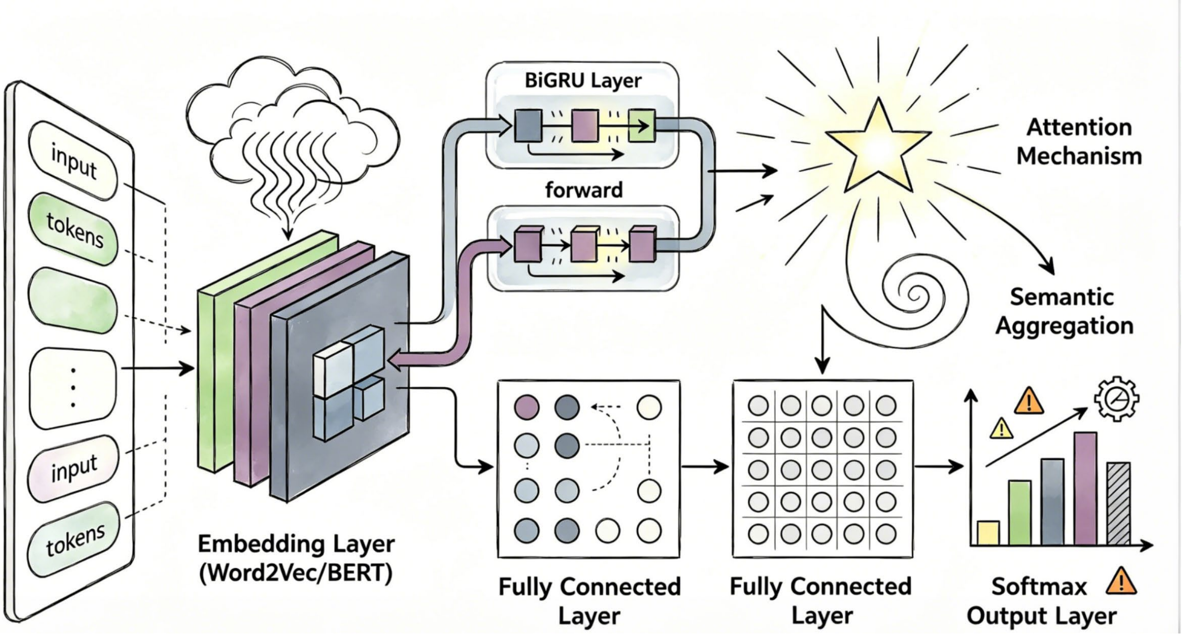


Figure 2. Architecture of the BiGRU-Attention fault classification model

To promote interpretability and facilitate engineering diagnostics, Figure 3 provides further detail on the data flow and information gating inside a BiGRU unit and the subsequent attention weighting, highlighting the seamless integration between temporal memory and adaptive focus at the token level.

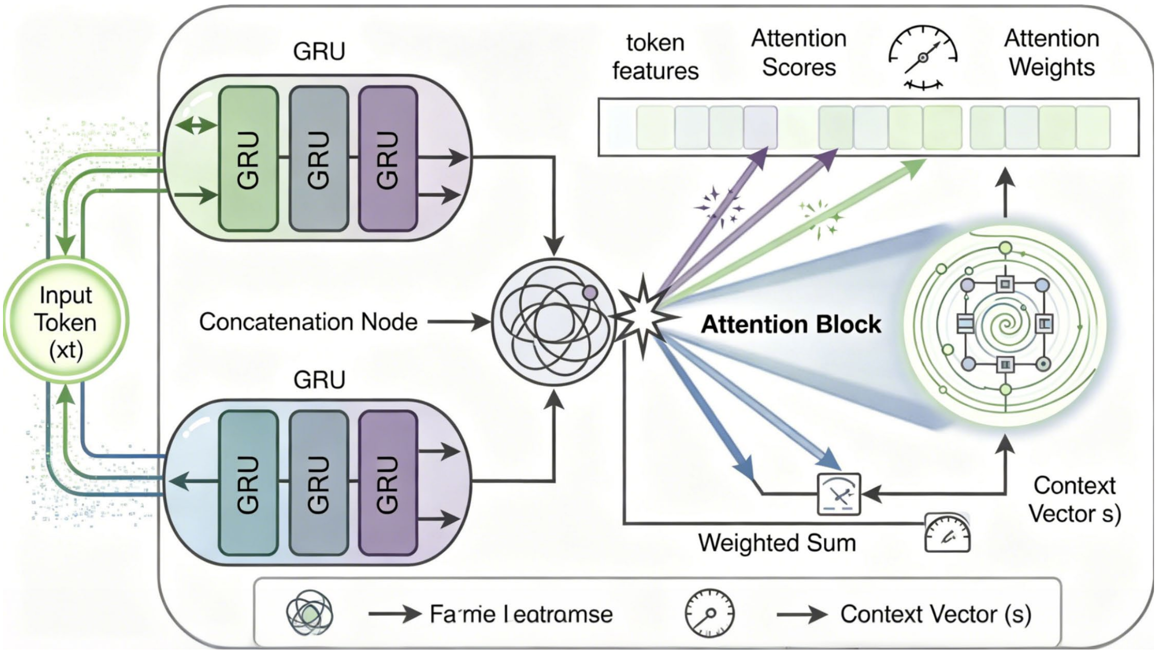


Figure 3. Structure of BiGRU unit and attention mechanism.

Training and Implementation Details

The BiGRU-Attention model's training and application processes are designed to ensure stability and repeatability in industrial text classification. When training a model, categorical cross-entropy is used as the loss function to calculate the degree to which the predicted class probabilities differ from the actual class labels. An L2-regularization term is introduced to the loss function for all trainable parameters in order to lessen overfitting and encourage more straightforward and broadly applicable models.

Adam is a common option for the optimisation method to handle non-stationary objectives and works well with adaptive-rate learning. When the validation loss ceases to decrease during training, the initial learning rate of 0.001 will be adjusted. Mini-batch training will be employed, with a batch size of 64 chosen for both gradient stability and computational ease. If the validation loss does not drop for ten consecutive epochs, early halting is triggered; otherwise, training will proceed to prevent overfitting.

grid search and empirical verification for systematic hyperparameter adjustment. For Word2Vec representations, the embedding layer is 200, while for BERT-based embeddings, it is 768. Each BiGRU direction has a hidden size of 128, and model correctness and computation cost must be balanced. In order to reduce overfitting in the face of noise and class imbalance in industrial data, a dropout rate of 0.5 is applied prior to the final fully connected layer.

To facilitate large-scale studies and expedite training, all computational experiments are carried out on high-performance GPU workstations. PyTorch is a high-performance, extensible, and user-friendly deep learning framework that offers numerous top-notch deep learning capabilities. For reproducibility, the training and preprocessing procedures are containerised using Docker.

Experimental Evaluation and Analysis

Dataset and Evaluation Protocol

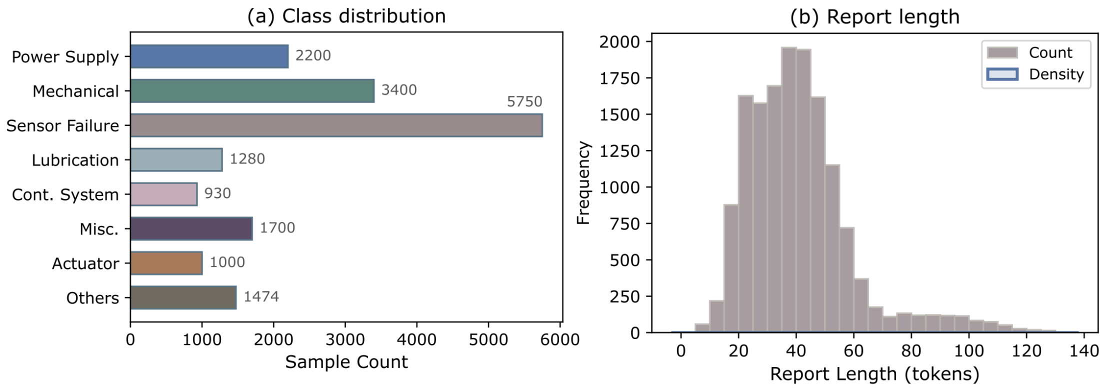
Experiments on the BiGRU-Attention model have been conducted using comprehensive data from the fault management systems of the two large-scale automotive manufacturing bases. Over the course of two years, there are 18,734 sets of annotated fault reports, each of which relates to a specific piece of equipment's malfunction and the corrective actions performed by trained maintenance staff. The dataset has various category distributions and text characteristics since it depicts various real-world working situations.

The eight primary types of faults—power supply anomalies, mechanical misalignment, sensor failures, lubrication issues, control system malfunctions, etc.—are dispersed unevenly among the classes. The rare classes of "Control System Fault" (5%) and "Lubrication Issue" (7%) are much smaller than the largest category, "Sensor Failure" (31%), as seen in Figure 4a. Together, the three most common defect categories make up over 63% of all records, which is consistent with the "long-tail" distribution of industrial reliability statistics. As a result, during operation, the classifier will have a poor detection rate for rare but extremely dangerous failure scenarios.

Figure 4b summarises the reports' textual characteristics and displays the token-based length distribution. Very long reports are uncommon; typically, they surpass 120 tokens and only happen in intricate, multi-step remediation scenarios. The median length of the reports is 39 tokens, and the modal range is 25–60 tokens. Attention mechanisms that can learn both short-form keywords and long-term, rich contexts are necessary because sequence models face challenges due to the varying lengths and descriptive levels of the sequences.

The inter-annotator agreement (Cohen's Kappa) reached roughly 0.93 as a result of the implementation of two layers of annotation quality assurance: first, in-plant technical staff carried out initial labelling, and second, senior engineers did an independent assessment. 13,113 training samples (70%), 1,873 validation samples (10%), and 3,748 test samples (20%) are the proportions that arise from data partitioning using a stratified approach. To minimise sampling-induced drift and approximate deployment conditions, the class proportions will be the same across all splits.

Figure 4c displays the frequency of out-of-vocabulary (OOV) technical phrases following domain-adapted tokenisation, along with the distribution of categorical labels and report length. The diversity and changes in the technical environment of industrial manufacture are indicated by a significant number of out-of-vocabulary (OOV) words in the low-frequency category; this presents an opportunity and a challenge for embedding-based models.



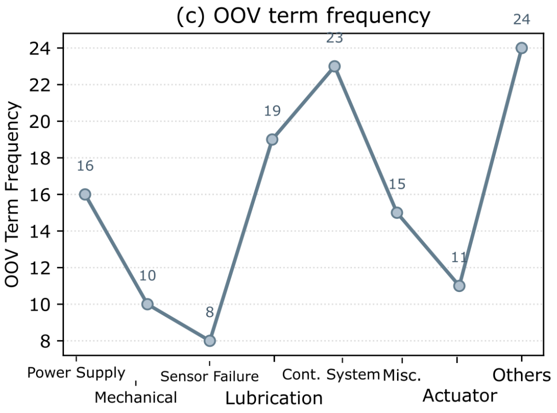


Figure 4. Dataset characteristics of industrial fault reports. (a) Class distribution. (b) Report length distribution. (c) OOV term frequency by category.

Results and Comparative Study

In order to compare the overall effectiveness of the new BiGRU-Attention model with both older and more recent base lines, a number of sets of tests have been carried out. SVM using TF-IDF features, LSTM, TextCNN, and refined BERT are the comparing models. The identical data split and feature engineering pipeline were used for training and testing each method so that the outcomes could be directly compared.

The BiGRU-Attention model achieved an accuracy of 91.5%, surpassing all deep learning baselines (BERT: 90.2%, LSTM: 88.7%, TextCNN: 87.9%) and the conventional SVM technique (82.4%). The quantitative test results are displayed in Table 1. BiGRU-Attention's macro-averaged precision, recall, and F1-score values are all comparatively excellent; in fact, the F1-score (91.1%) and recall (91.0%) are both high, indicating that it works well for both frequent and rare fault categories.

For a thorough visual comparison, Figure 5 displays the following three subplots. The four primary evaluation metrics for each model in this study—accuracy, precision, recall, and F1-score—are displayed in a grouped bar chart in Figure 5a. BiGRU-Attention shows comparatively large variations in F1 and recall while maintaining the leading values throughout. The minimum per-class recall is shown in Figure 5b; conventional models perform comparatively poorly when it comes to identifying uncommon but significant fault kinds. While both SVM and TextCNN have fallen below 0.70, BiGRU-Attention has maintained the minimum recall value above 0.86. The three models' macro-averaged AUC values are displayed in Figure 5c. It is evident that BiGRU-Attention has outperformed BERT (0.956) and LSTM (0.937) in terms of discriminative ability, with an AUC of 0.963.

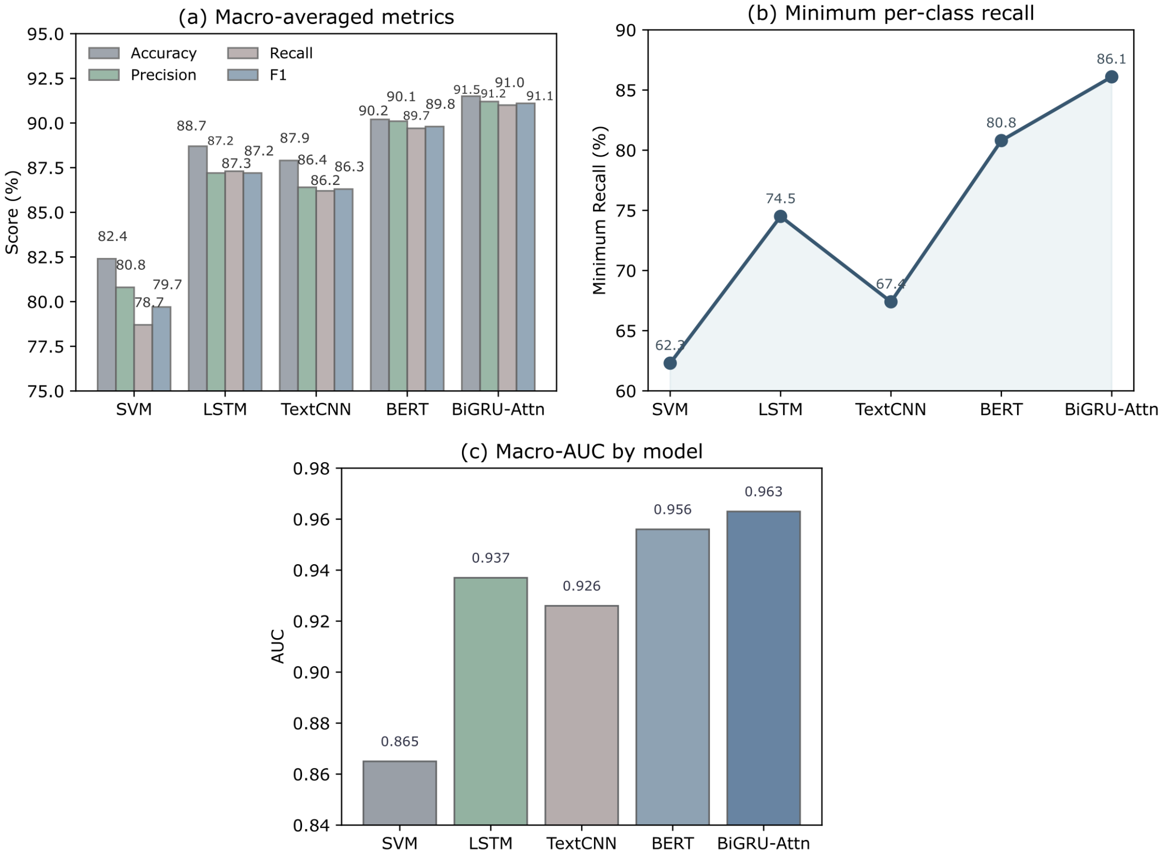


Figure 5. Comparative evaluation of model performance on the test set: (a) Macro-averaged metrics across models; (b) Minimum per-class recall comparison; (c) Area under ROC curve (AUC) for each model

Figure 6 also displays ROC curve analysis. The ROC curves for the three most common categories are displayed in Figure 6a. At all matching thresholds, BiGRU-Attention has a substantially higher true positive rate. The ROC curve graph for unusual categories in Figure 6b demonstrates that BiGRU-Attention is superior at separating these hard-to-distinguish errors. Lastly, a close-up of the region close to the origin is displayed in Figure 6c. This region is stable under low false-positive conditions, making it appropriate for early fault detection in an industrial setting.

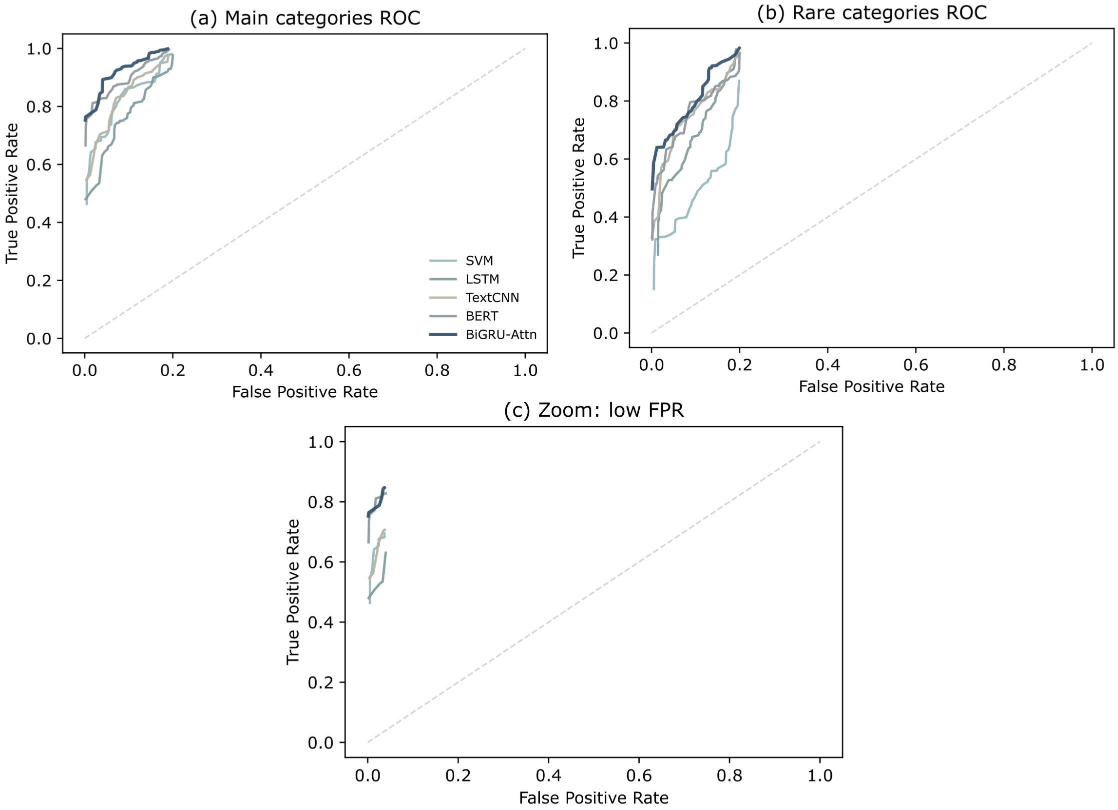


Figure 6. ROC curve analysis: (a) ROC curves for main categories; (b) ROC curves for rare categories; (c) Low false-positive region for critical scenarios.

Additional performance data is displayed by class. The suggested model maintains a low false alarm rate and a high recall rate for the minority classes of "Control System Fault" and "Lubrication Issue" in order to solve the issue of complicated and unbalanced industrial data. SVM and CNN-based models are not appropriate for field application because they are prone to misclassify fewer common classes, resulting in a significant decline in recall.

Ablation, Hyperparameter, and Error Analysis

For various reasons, numerous experiments have been carried out in succession in recent years to investigate the overall system's interpretability and general stability. The findings will improve our comprehension of the model's operation and provide more precise guidance for optimisation during the industrial deployment phase. The ablation study's Figure 7a illustrates how the primary architectural elements' individual contributions are measured. In particular, all major indications decrease when the attention module, the bidirectional mechanism, and the pre-trained embedding layer are removed one after the other [31]. It can be inferred that the attention mechanism aids in identifying and highlighting significant semantic cues from a variety of fault-report narratives because its absence results in the biggest decline in memory and F1-score [32]. For discriminative feature extraction, both past and future contexts in textual descriptions must be taken into account because the first issue with bidirectionality is a decrease in precision. A non-pre-trained (random) embedding is employed, as seen in Figure 7a; as a result, the model performs poorly and has to be initialised with domain-aware embeddings for industrial text.

The sensitivity of hyperparameters to variations in the validation F1-score is detailed in Figure 7b. Here, a moderate capacity and an optimised dropout (0.3) are near the ideal balance for generalisation and overfitting risk when the BiGRU hidden units, attention dimensionality, and dropout ratios are systematically varied [33]. For high-reliability applications, hyperparameter values must be relatively small since, as the figure illustrates, excessively large values will result in a significant drop in performance, even though the model is stable under tiny perturbations [34].

We further investigate the models' performance on the subgroup datasets under various operational settings, such as records with dense domain-specific jargon, reports that are either short or partial, and severely imbalanced classes, as illustrated in Figure 7c. The aforementioned investigation shows that BiGRU-Attention greatly outperforms the base model in the long-tail category and noisy environment; nevertheless, performance on extremely brief summaries is still being optimised. The aforementioned indicates that concentrating on data augmentation and model optimisation is the next phase [35].

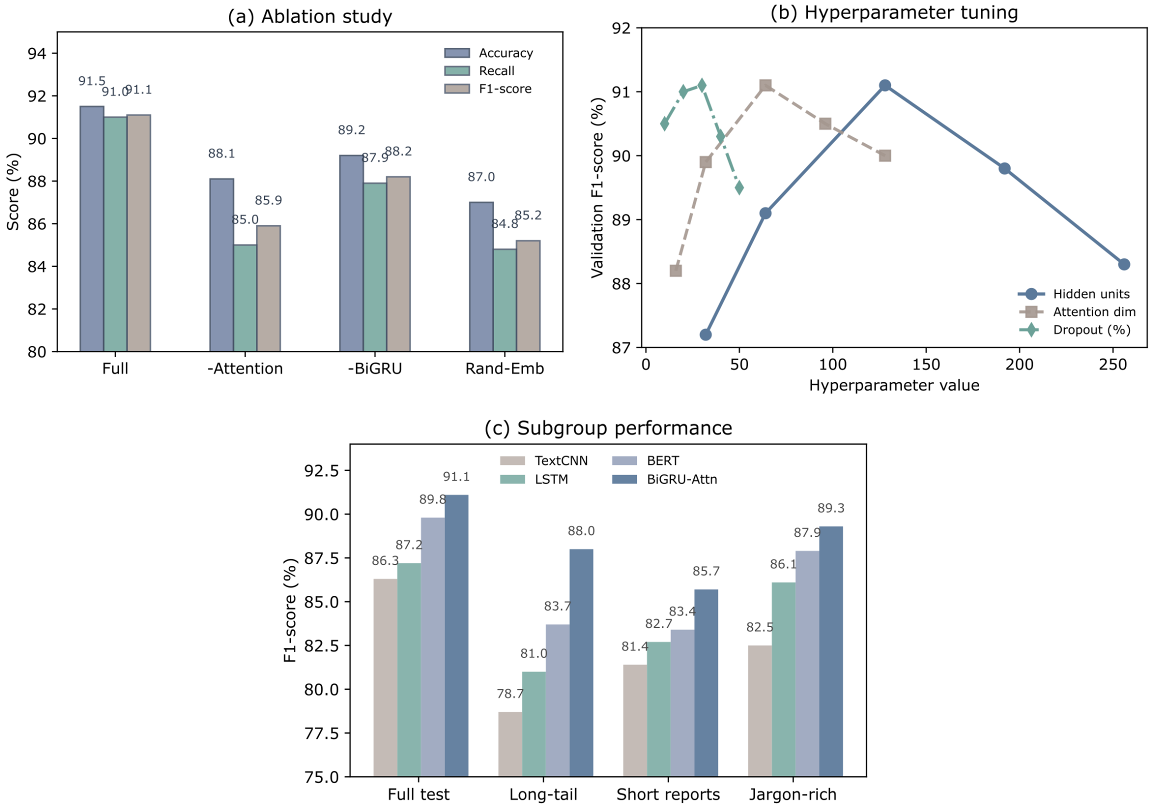
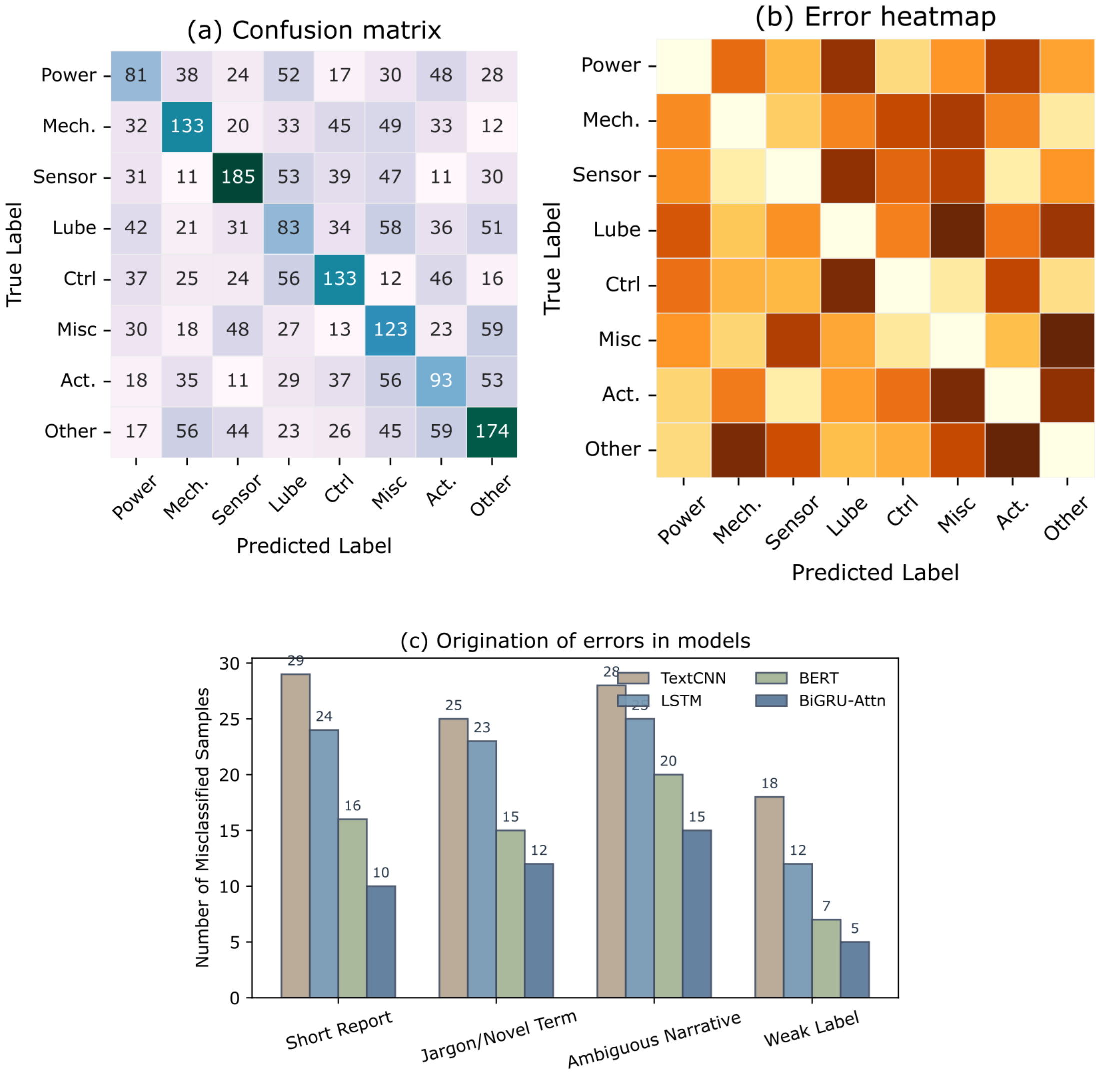


Figure 7. Model architecture and hyperparameter analysis: (a) Ablation study of model components; (b) Sensitivity analysis of key hyperparameters; (c) Subgroup performance evaluation across data environments.

The error diagnosis is displayed in Figure 8. The model is reasonably reliable because, as Figure 8a illustrates, the diagonal dominance of the confusion matrix is relatively high. Nevertheless, there are some clear clusters of off-diagonal values that indicate persistent misclassification of some neighbouring categories, such as "Sensor" and "Actuator" or between "Control System" and "Other." Semantic overlap and the intrinsic ambiguity of natural language maintenance data are to blame for this. More information about the error distribution is provided in Figure 8b. The heatmap indicates that the majority of misclassifications are still concentrated close to the true class, making them comparatively stable under a variety of industrial conditions.

The percentage of misclassification explanations for numerous models is detailed in Figure 8c. Here, TextCNN, LSTM, BERT, and BiGRU-Attention are compared in terms of the number of inaccurate predictions resulting from various reasons. In particular, "Short Report" is the primary cause of errors for the conventional models of TextCNN (29 errors) and LSTM (24 errors), while BiGRU-Attention (10 errors) has greatly decreased this. In a similar vein, BiGRU-Attention consistently has smaller "Jargon/Novel Term" and "Ambiguous Narrative" mistakes (12 and 15, respectively) than the other models; while BERT also exhibits some improvement, it falls short of the suggested model. Additionally, BiGRU-Attention has five times less errors due to a "weak label" than TextCNN (18 times), LSTM (12 times), and BERT (7 times).



	Power	Mech.	Sensor	Lube	Ctrl	Misc	Act.	Other
Power	81	38	24	52	17	30	48	28
Mech.	32	133	20	33	45	49	33	12
Sensor	31	11	185	53	39	47	11	30
Lube	42	21	31	83	34	58	36	51
Ctrl	37	25	24	56	133	12	46	16
Misc	30	18	48	27	13	123	23	59
Act.	18	35	11	29	37	56	93	53
Other	17	56	44	23	26	45	59	174

Figure 8. Error analysis visualizations: (a) Confusion matrix; (b) Error heatmap by category; (c) Typical misclassification case with detailed log context.

Conclusion

This research presents a BiGRU-Attention-based framework for industrial defect text categorisation that exhibits improved empirical performance over both conventional and sophisticated baselines, as well as structural innovation. The model's capacity to extract domain-specific linguistic information is improved by the use of adaptive attention mechanisms and bidirectional gated recurrent units. The aforementioned comparisons demonstrate that the suggested approach typically outperforms SVM, LSTM, TextCNN, and other well-known pre-trained models in terms of precision and recall, particularly in difficult scenarios like long-tail fault types and records with a lot of jargon. The findings above demonstrate how our architecture has improved the model's accuracy and made it suitable for industrial use.

The current model is quite stable under all typical operating settings, although it will make occasional mistakes since it is too sensitive to very brief or ambiguous maintenance data. The approach also requires the availability of annotated data and a well-defined label space; otherwise, it could be challenging to expand and adapt to various fault taxonomies at different times or across different businesses. Its computational efficiency has not satisfied the need for a large-scale real-time data stream and low-latency requirements in large-scale industrial systems, despite the fact that it is generally appropriate for all types of applications.

The model will be expanded in the future to accommodate multilingual and cross-domain text based on the aforementioned issues, hence enhancing interoperability across global industrial sectors. To lessen uncertainty and enhance the processing of uncommon or novel fault patterns, external domain knowledge—such as knowledge graphs and ontological resources—will be incorporated. To improve real-time adaptability and low-resource deployment, we will also investigate high-efficiency model compression technologies and adaptive learning methodologies. In the new era of industrial informatics, the aforementioned will further encourage the use and stability of intelligent text classification systems.

Author Contributions

Ghanim Al-Kuwaiti and Khalid Al-Shammari contribute to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Fahad Al-Dosari contributes to conceptualization, methodology, software. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Ruan, H., Dorneanu, B., Arellano-Garcia, H., Xiao, P., & Zhang, L. (2022). Deep learning-based fault prediction in wireless sensor network embedded cyber-physical systems for industrial processes. Ieee Access, 10, 10867-10879. 10.1109/ACCESS.2022.3144333

[2] Cheng, H., Luo, J., & Zhang, X. (2025). Multimodal industrial anomaly detection via uni-modal and cross-modal fusion. IEEE Transactions on Industrial Informatics. https://doi.org/10.1109/TII.2025.3552723

[3] Zhang, F., Liu, Q., Liu, Y., Tong, N., Chen, S., & Zhang, C. (2020). Novel fault location method for power systems based on attention mechanism and double structure GRU neural network. IEEE Access, 8, 75237-75248. https://doi.org/10.1109/ACCESS.2020.2988909

[4] Kong, F., Li, J., Jiang, B., Wang, H., & Song, H. (2021). Integrated generative model for industrial anomaly detection via bidirectional LSTM and attention mechanism. IEEE Transactions on Industrial Informatics, 19(1), 541-550. https://doi.org/10.1109/TII.2021.3078192

[5] Wang, B., Baraldi, P., & Zio, E. (2022). Deep multiadversarial conditional domain adaptation networks for fault diagnostics of industrial equipment. IEEE Transactions on Industrial Informatics, 19(8), 8841-8851. https://doi.org/10.1109/TII.2022.3222400

[6] Chi, W. W., Tang, T. Y., Salleh, N. M., Mukred, M., AlSalman, H., & Zohaib, M. (2024). Data augmentation with semantic enrichment for deep learning invoice text classification. IEEE Access, 12, 57326-57344. https://doi.org/10.1109/ACCESS.2024.3387860

[7] Zolanvari, M., Teixeira, M. A., & Jain, R. (2018, November). Effect of imbalanced datasets on security of industrial IoT using machine learning. In 2018 IEEE international conference on intelligence and security informatics (ISI) (pp. 112-117). IEEE. https://doi.org/10.1109/ISI.2018.8587389

[8] Saleh, R. A., & Metin Ertunç, H. (2024). Explainable attention-based fused convolutional neural network (XAFCNN) for tire defect detection: an industrial case study. Engineering Research Express, 6(1), 015090. https://doi.org/10.1088/2631-8695/ad23c8

[9] Zhao, Z., Zhang, Q., Yu, X., Sun, C., Wang, S., Yan, R., & Chen, X. (2021). Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study. IEEE Transactions on Instrumentation and Measurement, 70, 1-28. https://doi.org/10.1109/TIM.2021.3116309

[10] Ren, L., Meng, Z., Wang, X., Lu, R., & Yang, L. T. (2020). A wide-deep-sequence model-based quality prediction method in industrial process analysis. IEEE Transactions on Neural Networks and Learning Systems, 31(9), 3721-3731. https://doi.org/10.1109/TNNLS.2020.3001602

[11] Salur, M. U., & Aydin, I. (2020). A novel hybrid deep learning model for sentiment classification. IEEE Access, 8, 58080-58093. https://doi.org/10.1109/ACCESS.2020.2982538

[12] Kidambi Raju, S., Ramaswamy, S., Eid, M. M., Gopalan, S., Alhussan, A. A., Sukumar, A., & Khafaga, D. S. (2023). Enhanced dual convolutional neural network model using explainable artificial intelligence of fault prioritization for industrial 4.0. Sensors, 23(15), 7011. https://doi.org/10.3390/s23157011

[13] Hu, W., Yang, G., Li, Y., Cao, W., & Wu, M. (2023). Root cause identification of industrial alarm floods using word embedding and few-shot learning. IEEE Transactions on Industrial Informatics, 20(2), 1465-1475. https://doi.org/10.1109/TII.2023.3274223

[14] Li, D., Zhang, H., An, L., Dong, J., & Peng, K. (2024). A quality-related fault detection method based on deep spatial–temporal networks and canonical variable analysis for industrial processes. IEEE Sensors Journal, 24(18), 29047-29055. https://doi.org/10.1109/JSEN.2024.3435013

[15] Abid, F. B., Sallem, M., & Braham, A. (2019). Robust interpretable deep learning for intelligent fault diagnosis of induction motors. IEEE Transactions on Instrumentation and Measurement, 69(6), 3506-3515. https://doi.org/10.1109/TIM.2019.2932162

[16] Cao, Q., Samet, A., Zanni-Merk, C., de Bertrand de Beuvron, F., & Reich, C. (2020). Combining chronicle mining and semantics for predictive maintenance in manufacturing processes. Semantic Web, 11(6), 927-948. https://doi.org/10.3233/SW-200406

[17] Jang, K., Pilario, K. E. S., Lee, N., Moon, I., & Na, J. (2023). Explainable artificial intelligence for fault diagnosis of industrial processes. IEEE Transactions on Industrial Informatics, 21(1), 4-11. https://doi.org/10.1109/TII.2023.3240601

[18] Wang, X., Yu, J., & Fong, S. (2023, November). Enhancing Official Policy Document Classification with BERT and Keyword Augmentation. In 2023 Third International Conference on Digital Data Processing (DDP) (pp. 134-142). IEEE. https://doi.org/10.1109/DDP60485.2023.00035

[19] Bao, G., Zhang, H., Wei, Y., Gu, D., & Liu, S. (2020). Fault diagnosis of reciprocating compressor based on group self-attention network. Measurement Science and Technology, 31(6), 065014. https://doi.org/10.1088/1361-6501/ab7280

[20] Cui, Q., El-Arroudi, K., & Weng, Y. (2019). A feature selection method for high impedance fault detection. IEEE Transactions on Power Delivery, 34(3), 1203-1215. https://doi.org/10.1109/TPWRD.2019.2901634

[21] Wu, Y., Guo, D., Liu, H., & Huang, Y. (2020). An end-to-end learning method for industrial defect detection. Assembly Automation, 40(1), 31-39. https://doi.org/10.1108/AA-08-2018-114

[22] Yin, Z., Zhang, X., Song, Z., & Ge, Z. (2023). Adversarial learning from imbalanced data: A robust industrial fault classification method. IEEE Transactions on Information Forensics and Security, 19, 1870-1882. https://doi.org/10.1109/TIFS.2023.3343073

[23] De Luca, R., Ferraro, A., Galli, A., Gallo, M., Moscato, V., & Sperli, G. (2023). A deep attention-based approach for predictive maintenance applications in IoT scenarios. Journal of Manufacturing Technology Management, 34(4), 535-556. https://doi.org/10.1108/JMTM-02-2022-0093

[24] Zhang, Q., Zheng, R., Zhao, Z., Chai, B., & Li, J. (2020, April). A textcnn based approach for multi-label text classification of power fault data. In 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA) (pp. 179-183). IEEE. https://doi.org/10.1109/ICCCBDA49378.2020.9095584

[25] Chen, H., Luo, H., Huang, B., Jiang, B., & Kaynak, O. (2023). Transfer learning-motivated intelligent fault diagnosis designs: A survey, insights, and perspectives. IEEE Transactions on Neural Networks and Learning Systems, 35(3), 2969-2983. https://doi.org/10.1109/TNNLS.2023.3290974

[26] Zhang, Y., Feng, K., Ma, H., Yu, K., Ren, Z., & Liu, Z. (2022). MMFNet: Multisensor data and multiscale feature fusion model for intelligent cross-domain machinery fault diagnosis. IEEE Transactions on Instrumentation and Measurement, 71, 1-11. https://doi.org/10.1109/TIM.2022.3213016

[27] Han, H., Wang, J., Wang, X., & Chen, S. (2022). Construction and evolution of fault diagnosis knowledge graph in industrial process. IEEE Transactions on Instrumentation and Measurement, 71, 1-12. https://doi.org/0.1109/TIM.2022.3200429

[28] Lei, M., Zhou, Y., Zhou, L., Zheng, J., Li, M., & Zou, L. (2019). Noise-robust wagon text extraction based on defect-restore generative adversarial network. IEEE Access, 7, 168236-168246. https://doi.org/10.1109/ACCESS.2019.2954475

[29] Chen, J., Tang, J., & Li, W. (2023). Industrial edge intelligence: Federated-meta learning framework for few-shot fault diagnosis. IEEE Transactions on Network Science and Engineering, 10(6), 3561-3573. https://doi.org/10.1109/TNSE.2023.3266942

[30] Liu, P., Wang, X., Xiang, C., & Meng, W. (2020, August). A survey of text data augmentation. In 2020 International Conference on Computer Communication and Network Security (CCNS) (pp. 191-195). IEEE. https://doi.org/10.1109/CCNS50731.2020.00049

[31] Bosio, A., Bernardi, P., Ruospo, A., & Sanchez, E. (2019, March). A reliability analysis of a deep neural network. In 2019 IEEE Latin American Test Symposium (LATS) (pp. 1-6). IEEE. https://doi.org/10.1109/LATW.2019.8704548

[32] Li, Y., Zhou, Z., Sun, C., Chen, X., & Yan, R. (2022). Variational attention-based interpretable transformer network for rotary machine fault diagnosis. IEEE transactions on neural networks and learning systems, 35(5), 6180-6193. https://doi.org/10.1109/TNNLS.2022.3202234

[33] Srivastava, S., Bala, N., Gupta, A., Priya, B. D., Kumar, S., & Raj, A. (2024, December). Optimization of sentiment analysis models using bayesian hyperparameter tuning. In 2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA) (pp. 1-6). IEEE. https://doi.org/10.1109/ICAIQSA64000.2024.10882347

[34] Iqbal, R., Maniak, T., Doctor, F., & Karyotis, C. (2019). Fault detection and isolation in industrial processes using deep learning approaches. IEEE Transactions on Industrial Informatics, 15(5), 3077-3084. https://doi.org/10.1109/TII.2019.2902274

[35] Zhu, L., Tian, N., Li, W., & Yang, J. (2022). A text classification algorithm for power equipment defects based on random forest. International Journal of Reliability, Quality and Safety Engineering, 29(05), 2240001. https://doi.org/10.1142/S0218539322400010