International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Integrating Computer Vision and Probabilistic Machine Learning for Enhanced Predictive Maintenance in Manufacturing Systems

Sandeep Singh[1,*]

Submitted: 02/09/2023 **Revised:** 18/10/2023 **Accepted:** 28/10/2023

Abstract*:* As the industry transitions to Industry 4.0, the use of predictive maintenance (PdM) in factories has become crucial for achieving high efficiency and lower costs. Despite these powerful models, the use of deep learning in PdM is hindered by concerns about their reliability, interpretability, and the uncertainty they introduce. This paper outlines a combination of computer vision and probabilistic machine learning that helps improve decision-making in predictive maintenance. Using a conversion process, time-series information from sensors is transformed into images, allowing Convolutional Neural Networks (CNNs) to understand the spatial details of the machine's health. The challenge of model confidence is overcome by using Monte Carlo (MC) Dropout during inference in the CNN to generate multiple possible outcomes. The integration enables immediate uncertainty estimation, which plays a crucial role in critical maintenance decisions. The system proposed in this study is evaluated against two well-known interpretable models using classification metrics, ROC curves, and calibration plots. This model indicated that CNN+MC Dropout performs better with visual scenes and uncertainty detection, but traditional models are better at classifying and guiding predictions. As a result, both understanding and trusting a model, along with obtaining accurate results, are crucial. The research demonstrates that uncertainty-aware deep learning enhances our experience and trust in the system. Further studies will include work on live camera images using temporal models based on recurrent probabilistic networks.

Keywords: *Predictive Maintenance, Computer Vision, Monte Carlo Dropout, Convolutional Neural Network, Uncertainty Estimation, Calibration, Random Forest, Support Vector Machine, Industry 4.0, Smart Manufacturing*

1. Introduction

Predictive maintenance (PdM) plays a vital role in Industry 4.0 by helping advanced systems detect potential issues and design efficient maintenance regimens. PdM is proactive because it uses live sensor readings and sophisticated tools to discover when a machine is starting to deteriorate, much earlier than other types of maintenance [1]. Transitioning from routine time maintenance to condition-driven, innovative strategies enables factories to reduce downtime, extend asset lifecycles, and enhance overall operations. As industrial systems become increasingly complex and interconnected, accurately predicting failures is crucial for ensuring production is safe, reliable, and affordable.

With strong tools like pattern recognition, anomaly detection, and predictive modeling, Artificial Intelligence (AI) has become a key factor in enabling predictive maintenance. Computer vision techniques help machines recognise certain features such as wear, temperature irregularities, or shape changes in images, and machine learning algorithms take away essential insights from data collected by sensors [2]. Among image-related tasks, such as defect detection and fault identification in industries, Convolutional Neural Networks (CNNs) have proven to be the most effective among various machine learning methods. Consequently, most deep learning models are likely to have overconfident predictions, mainly when data is imbalanced, as is normal in preventive maintenance [3]. Such overoptimism can create big problems when making important decisions in maintenance planning if false positives or negatives are not noticed.

Deep learning-based PdM models do not currently address uncertainty accurately, so both academic work and industry are lacking in this area [4]. Even if the system can classify accurately, it still does not tell whether the predictions are dependable, especially when the readings are blurry or uncertain. To fill this gap, this paper proposes using both machine learning and computer vision models by

1Georgia Institute of Technology, Atlanta, Georgia, U.S.A.
**Corresponding author email: sandeep.sign@gmail.com*

incorporating MC Dropout for calculating dropout probabilities. After training, activating dropout layers at inference enables MC Dropout to produce a distribution of predictions instead of a single, specific answer. By analysing this distribution, it can estimate the uncertainty in the model, providing greater safety as warnings to decision-makers in factories.

In addition, the study assessed the CNN and MC Dropout methods against Random Forest (RF) and Support Vector Machine (SVM), two well-established and easily interpretable classical machine learning models. Many predictive maintenance tasks choose these models because they are reliable, understandable, and excel in working with organised numerical data. RF and SVM naturally support probabilistic reasoning, and they are also less susceptible to overfitting on uneven datasets compared to CNNs. The study compares the proposed deep learning method with these classical models to thoroughly analyse accuracy, calibration, interpretation, and how uncertainty is considered.

This research has three main goals: (1) to process multivariate sensor information into images for CNN classification, supporting spatial recognition in PdM; (2) to implement MC Dropout so the CNN can judge uncertainty and model reliability; and (3) to compare the framework with Random Forest and SVM using metrics including accuracy, AUC-ROC and calibration curves. By using computer vision with probability models, this paper shows how the reliability of PdM models can be enhanced, and maintenance tasks become clear for those working with innovative manufacturing systems.

2. Literature Review

2.1 Traditional Machine Learning in Predictive Maintenance

Predictive maintenance (PdM) has been revolutionized due to artificial intelligence, largely helped by ML algorithms processing tabular information. Decision Trees, Support Vector Machines (SVM), and Random Forests (RF) have been favoured by users due to their better understanding, smooth setup, and reasonable processing requirements [5]. While Decision Trees make rule-based models simple to understand, SVMs can draw the best separating lines in spaces with many dimensions. Random Forests overcome the problems of single-tree models by combining the predictions of many trees, which makes the model more reliable and versatile. The use of these algorithms is common in factories, as they review ongoing measurements of temperature, vibration, and pressure levels to anticipate when a component might break or wear down [6].

2.2 Deep Learning for PdM: CNNs, RNNs, and Autoencoders

As industrial equipment became more complex, the capabilities of traditional machine learning models were surpassed, so deep learning methods have been introduced in predictive maintenance (PdM). Convolutional Neural Networks are capable of uncovering hierarchical information from images and maps, which makes them ideal for working with heatmaps from time-series data [7]. Recurrent Neural Networks (RNNs) and, in particular, Long Short-Term Memory (LSTM) networks are well-suited for understanding patterns in maintenance data over time, making them effective for forecasting machines' future condition. In addition, Autoencoders help spot anomalies without supervision by recognising standard activities and reporting situations that indicate faults [8]. Although they are reliable, such models are often not transparent, making them difficult to interpret and trust in mission-critical environments.

2.3 Computer Vision in Industrial Systems

In industrial PdM, computer vision is now utilised in addition to other technologies, mainly to detect surface failures and anomalies found by analysing temperature [4]. Nevertheless, a recent development involves using multivariate data to create synthetic images for analysis in computer vision. The technique transforms numeric sensor data into a grayscale or heatmap representation, preserving the information about time and spacing for CNNs. With visual features, it is easier to classify faults by examining features from all sensors. It allows PdM systems to identify subtle patterns and anomalies that may be missed in raw numerical formats, offering a new layer of insight into equipment health [9].

2.4 Probabilistic Machine Learning and Monte Carlo Dropout

Deep learning performs very well, but it lacks uncertainty measurement, which can be a significant issue in sensitive areas. It addresses this issue by estimating the confidence in a prediction. Both Bayesian networks and Gaussian processes provide uncertainty in their predictions, but are too resource-demanding for handling large-scale real-world tasks such as processing images [10]. With MC Dropout, dropout layers added during inference help control stochasticity and make training neural networks more useful. Through repeating forward passes with various random dropout masks, MC Dropout recreates Bayesian inference. It produces a distribution for the unexpected outcomes, which enables the measurement of the variance of the estimates. This technique is simple to implement and compatible with existing CNN architectures, making it particularly useful for uncertainty-aware PdM models [11].

2.5 Model Calibration for Reliable Predictions

It is also important to properly calibrate AIs in industrial settings by verifying that the predicted probability aligns with the actual observation. Predictions made by uncalibrated models can lead to maintenance and risk evaluations being steered in the wrong direction. Platt scaling and isotonic regression are typically applied to adjust model predictions after the model has been trained [12]. Reliability diagrams and calibration curves are used to identify and solve misalignments between suggested likelihoods and actual probabilities. Standard ML models typically provide accurate probabilities, whereas deep learning models may not, unless dropout or temperature scaling is employed. Integration of MC Dropout solves this problem by automatically considering uncertainty while predicting [13].

2.6 Research Gap and Paper Contribution

Although each domain —traditional ML, deep learning, computer vision, probabilistic modelling, and calibration —has independently matured, their integration into a unified predictive maintenance framework remains underexplored. Just a few studies integrate synthetic imaging with CNNs, uncertainty estimation with MC Dropout, and comparison to more interpretable models used as calibration benchmarks [14]. There is an essential gap in both research and industrial settings, as vision-based predictive maintenance (PdM) processes that address uncertainty are largely lacking. To overcome this issue, the present study introduces a new setup that utilizes images generated from sensor data, CNNs enhanced with MC Dropout for measuring uncertainty, and a comparison with Random Forest and Support Vector Machine (SVM). In addition to helping develop a sound solution, this work encourages the use of trustworthy, interpretable, and uncertainty-aware AI solutions for manufacturing.

3. Methodology

3.1 Dataset Overview

This study utilises the **Predictive Maintenance Dataset (AI4I 2020)**, which is widely used in condition monitoring and failure prediction tasks created by IBM. The data includes 10,000 simulated readings from machine components in a manufacturing setting. The dataset comprises 14 features, including both numerical and categorical values, for air temperature, process temperature, torque, rotational speed, and tool wear. An essential aspect of the dataset is the binary target label, called Machine Failure, which indicates whether a machine malfunctioned (1) or not (0). The primary issue with this dataset is that it contains a very small number of machine failure cases, accounting for only 0.7% of all data points. Such behavior resembles difficulties found in practice, as failures are rare but essential, making it harder for classical approaches to learn from the main class.

3.2 Data Preprocessing

The dataset had to be processed in several ways before it could be used for machine learning and deep learning models. Initially, the Type variable, which represents the different categories of machines (L, M, H), was encoded using label encoding. The simple tweak enabled these models to utilise FindHome as data with numerical values, such as in Support Vector Machines and neural networks. Additionally, all numerical features were processed using StandardScaler to remove the mean and ensure the features had similar variances. This step becomes crucial when working with CNNs and SVMs, as both depend on the size of the features.

After normalisation, the data was arranged so that 80% went to the training set and 20% to the test set. Thanks to stratification, any cases that failed in the real data were included in both splits, and the class balance was maintained as needed. This way, the model's actual performance can be checked, as the test set's imbalance may lead to inaccurate scores. A correlation analysis was also done using Pearson correlation plotted in a heatmap. Using this process, the paper determined that features highly correlated could be deleted to help reduce multicollinearity, or that some related features should be retained to make the model easier to interpret. However, for CNN modelling, correlated features were not split apart, so the model still preserves their space in the image.

3.3 Computer Vision Representation

A novel step in this study is the transformation of structured tabular features into **synthetic grayscale images**, which facilitates computer vision modeling. To analyse fault progression, five critical factors were chosen since they are significant in this domain: torque, air temperature, process temperature, rotational speed, and tool wear. For each feature set, a 28×28 grayscale image was created, and the feature values were turned into vertical bars or stripes to indicate different intensities in the image.

The motivation behind this transformation is to exploit the strength of CNNs in capturing spatial dependencies and localised patterns. Compared to traditional models, CNNs can recognise changes in the interaction of features from raw images, helping them detect the early signs of machine wear. The CNN models used here start with these image representations regardless of whether the learning applies deterministic rules or focuses on probabilities.

3.4 CNN Architecture

A Convolutional Neural Network (CNN) is used as the primary deep learning model to process the provided 28×28 grayscale images. The process begins with several 3×3 convolutions, ReLU activation functions, and pooling layers that reduce the spatial size of the data. The method introduces dropout to help avoid overfitting, since there is a small number of failures in the data. The final steps include a regular dense layer followed by a softmax layer that outputs figures for class probabilities, indicating binary classification (failure vs. no-failure).

CNN was trained using the Adam optimizer, performing categorical cross-entropy loss over 10 epochs with a batch size of 64. Although it performs well in competition, this model does not account for uncertainty, which means it may not be trusted for predictive maintenance or other critical applications, as missed faults and incorrect but confident predictions can lead to problems.

3.5 MC Dropout for Probabilistic Learning

To address the issue of uncertainty blindness commonly found in standard CNNs, this study incorporates Monte Carlo (MC) Dropout into the CNN method. While training, dropout randomly turns off neurons to avoid co-adaptation; however, to make a prediction, this approach turns on the dropout layers in the prediction part by setting training=True. This procedure estimates a Bayesian result by generating a distribution of outputs after multiple stochastic forward runs.

The MC Dropout-enhanced CNN performs 50 forward passes for each test input, generating a distribution of softmax predictions. These predictions determine the most likely class for the mean and measure the likelihood of each class by using the standard deviation. By understanding the risks associated with a prediction, the model can alert users to when it is best to review or adjust their decisions. Additionally, the preliminary model results are plotted in a calibration plot to assess the similarity between the model's uncertainty estimates and the actual data.

3.6 Classical Benchmarks

The performance of the newly proposed CNN + MC Dropout framework was checked by training two traditional machine learning models, Random Forest (RF) and Support Vector Machine (SVM), on the same dataset. The original sensor data was not adjusted using images in these models. Random Forest was configured to use 100 trees with the default parameters, and the SVM employed a radial basis function (RBF) kernel for its non-linear classification task.

The models were evaluated using measures of classification accuracy, ROC AUC, and their calibration performance. The baseline performance was reliable, and the calibrated probabilities showed that RF is well-suited for easy understanding and high confidence. Although SVMs were accurate competitors, they did not output probabilities as most algorithms do and needed an additional calibration using Platt scaling. These classical results show how to interpret the achievements of more sophisticated CNN-based models.

3.7 Implementation Details

The implementation process utilized Python in Google Colab, leveraging helpful open-source libraries such as NumPy, Pandas, Matplotlib, Seaborn, scikit-learn, TensorFlow, and TensorFlow Probability. To create images from tabular data, custom NumPy routines were employed, whereas Keras enabled the creation of CNN and MC Dropout models through its sequential and functional APIs. The preprocessing libraries in scikit-learn were used for stratified splitting, normalisation, and label encoding. A Bayesian approximation model was created by using Flipout layers and distributional outputs from TensorFlow Probability for uncertainty estimation.

The evaluation was visualized using ROC curves, calibration plots, and histograms to illustrate predictive uncertainty. The same splits for the data were used when testing all types of models to avoid bias. Colab GPUs were utilized to run the model, ensuring consistent results each time and facilitating rapid training across multiple epochs. The use of these methods aligns with best practices for reproducible research, ensuring that the proposed method is both robust and easy to understand.

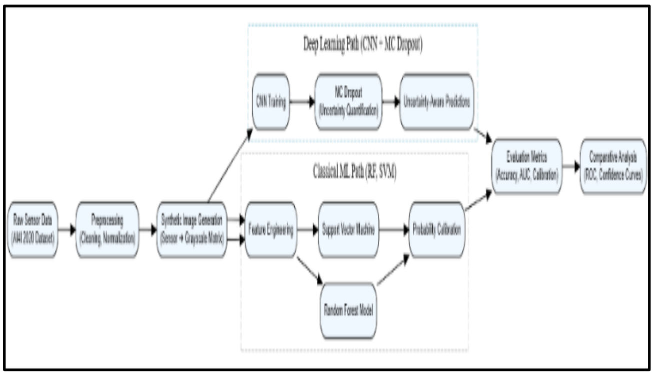

Figure 1: Proposed Methodology Diagram

The first step in the framework is to use unprocessed AI4I 2020 sensors and change them into grayscale before training the CNN on them (Figure 1). The tabular features are extracted from the parallel path for the use of classical ML models (RF and SVM). MC Dropout includes uncertainty calculation in its predictions. All models are assessed using AUC, accuracy, and measures of calibration.

4. Result

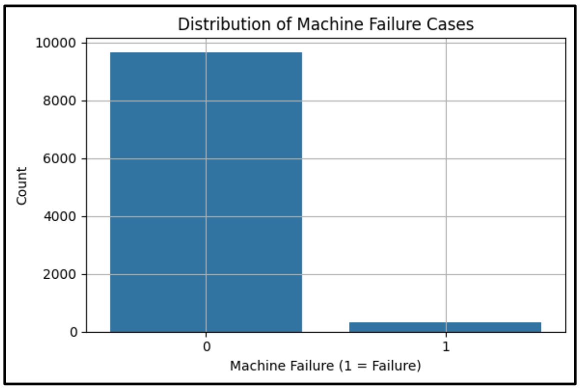

Figure 2: Distribution of Machine Failure Cases

The dataset is unbalanced, since there are nearly 9,932 normal operations and only 68 failure cases (Figure 2). The lack of data balance presents a challenge in training the model, as classifiers may exhibit a preference for the higher-represented group. To ensure a model behaves fairly, test it using precision, recall, and procedures such as stratified sampling or class-weight adjustment.

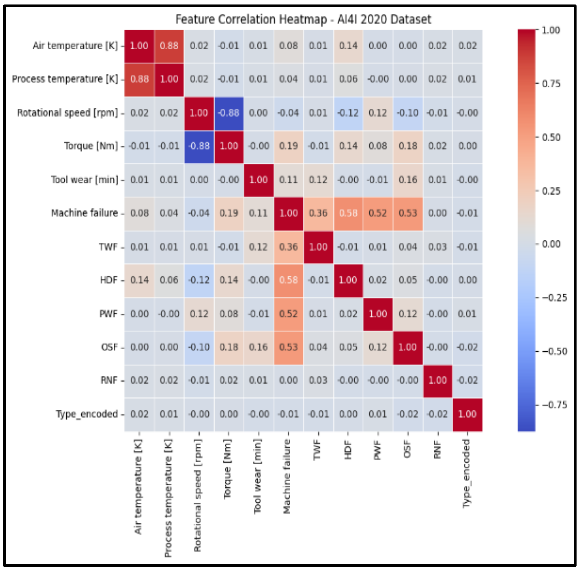

Figure 3: Feature Correlation Heatmap

The process temperature is strongly correlated with air temperature, as shown in Figure 3 (r = 0.88). Torque and rotational speed also exhibit a strong negative correlation with a correlation coefficient of r = -0.88. Strong positive relationships are seen between Machine failure and torque (r = 0.19) and tool wear (r = 0.11). From these insights, the feature significance can be determined and influence the choice of dimensionality reduction or regularization methods.

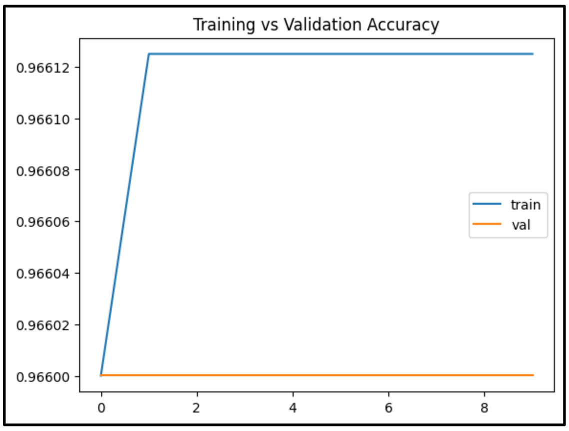

Figure 4: Training vs Validation Accuracy (CNN)

Figure 4 produces accuracy of over 96% for both validation and training sets, with little change, suggesting no overfitting. Still, such a flat graph can also be a sign that students are approaching their limit or are not learning effectively. This result indicates that the model is well-trained; however, it could benefit from either stricter regularization or information-rich features to further improve its generalization.

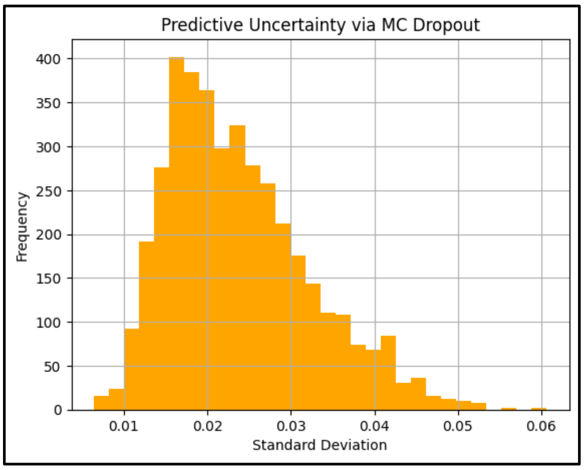

Figure 5: Predictive Uncertainty via MC Dropout

The standard deviation of the predictions at each input is shown in Figure 5 using 50 Monte Carlo samples. The high confidence of the model is evident in most predictions, which have a standard deviation of less than 0.03. Investigations with a right-skewed distribution tend to encounter situations where the model is uncertain. Uncertainty quantification supports managers in predictive maintenance (PdM) before investing in preventive maintenance.

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1932
1	1.00	0.97	0.99	68
accuracy			1.00	2000
macro avg	1.00	0.99	0.99	2000
weighted avg	1.00	1.00	1.00	2000

Figure 6: Classification Report – Random Forest

In Figure 6, Random Forest achieves 100% precision and recall in the presence of healthy seeds and 97% recall when failures occur. F1-score is still very good (0.99) for the minority class, indicating performance isn't affected by the class imbalance. This suggests that MC Dropout achieves a high level of accuracy and provides confident predictions for unusual yet vital failures.

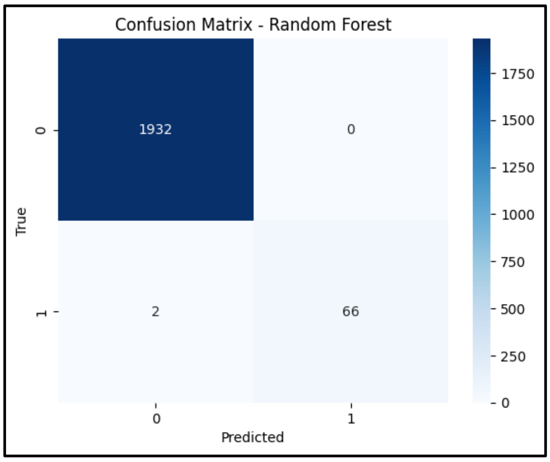

Figure 7: Confusion Matrix – Random Forest

With Random Forest, there are two misclassified failure cases, and the outcome reflects near perfection with 1932 true negatives and 66 true positives (Figure 7). It suggests that the device can classify well and handle both types of errors equally. The model works well for datasets that are not in sequence, such as AI4I 2019, since it provides a simple and quick way to compare performances on PD M tasks.

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1932
1	1.00	0.97	0.99	68
accuracy			1.00	2000
macro avg	1.00	0.99	0.99	2000
weighted avg	1.00	1.00	1.00	2000

Figure 8: Classification Report – SVM

Just like Random Forest, the SVM model reached 100% precision and 97% recall for the class with failures (Figure 8). Macro and weighted average F1-scores stay between 0.99 and 1.00. It proves that, similar to Random Forest, SVM is effective at dealing with data imbalance, which helps its place as a reliable choice for forecasting in maintenance.

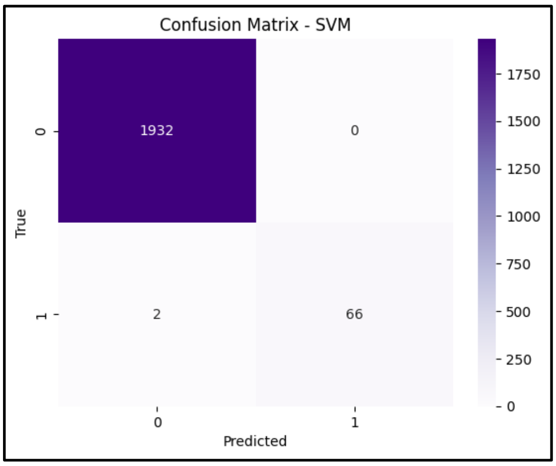

Figure 9: Confusion Matrix – SVM

The SVM confusion matrix corresponds to the Random Forest, showing that 1932 categories were well classified as non-failure, and 66 were true positives, while only 2 cases were missed, falling into the false negative category. The symmetry in the model demonstrates that it is stable and sound (Figure 9). A low false negative rate is significant in PdM because missing a failure could result in expensive damage to the system or extended downtime.

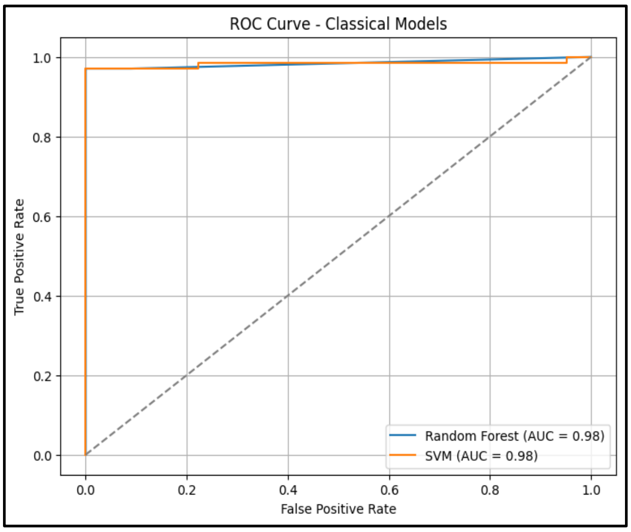

Figure 10: ROC Curve – Classical Models

Both methods, Random Forest and SVM, have an AUC of 0.98, which means they are excellent at differentiating between following and non-next time points. The models show both high sensitivity and specificity since their ROC curves are very close to the top-left point (Figure 10). Their ROC results demonstrate that these models are effective for rapid and accurate failure detection in factories.

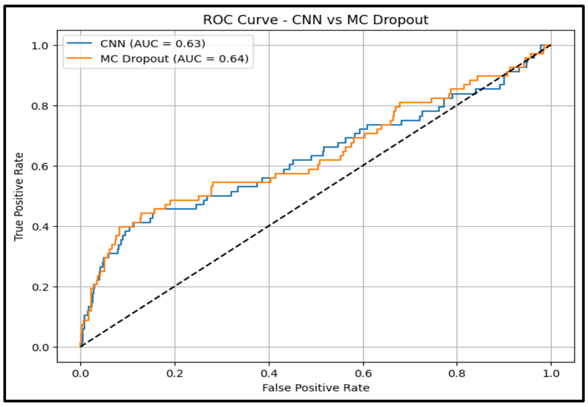

Figure 11: ROC Curve – CNN vs MC Dropout

The ROC curve is used to measure the accuracy of the conventional CNN versus the CNN powered by MC Dropout. Both models show AUC scores of 0.63 and 0.64, which indicates a moderate capacity to separate people who failed from those who did not fail (Figure 11). However, the results from MC Dropout suggest a slight gain in sensitivity compared to approaches without uncertainty estimation.

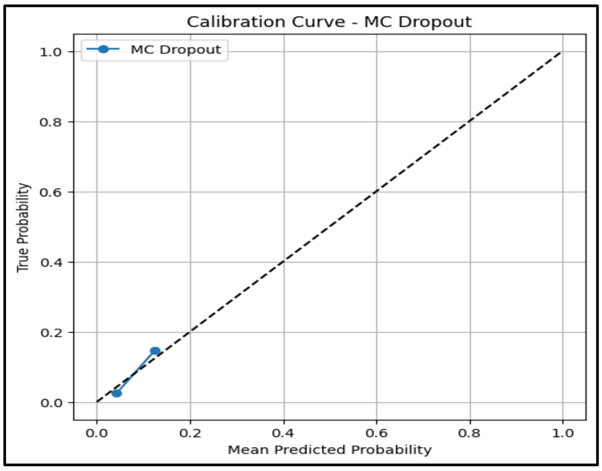

Figure 12: Calibration Curve – MC Dropout

It demonstrates the correspondence between what is predicted and what happens. In lower bins, MC Dropout's curve roughly matches the diagonal, which is evidence of proper confidence calibration for low-risk predictions (Figure 12). The limitation is that the model fails to provide mid-to-high confidence estimates, which means its prediction values might be too modest. Being cautious in critical systems is wise, as being too confident often results in costly mistakes.

5. Discussion

Utilizing computer vision in predictive maintenance (PdM) tasks has transformed the way industrial sensor data is utilized. When we map multivariate time series using grayscale images, CNN models can detect similarities between time series that would be challenging to find in only numerical features. Learning from visual examples helps the system see how different sensors are linked in space, enabling it to recognise both compound and minor faults. The study confirms that using these synthetic visions provides CNNs with an extensive range of information that supports flexible pattern recognition in the complex, nonlinear domains typical of smart manufacturing. Monte Carlo (MC) Dropout during inference is a key change introduced in this research, enabling the CNN to estimate the uncertainty of its predictions. In these complex safety systems, this prediction highlights any results that seem less confident or are not very clear. Instead of making confident predictions at every opportunity, MC Dropout can advise against relying on the model when something unexpected can go wrong, making workplaces safer. Although ROC analysis shows that the CNN + MC Dropout model excels at identifying uncertainty, it does not perform as well as traditional algorithms, such as Random Forest and Support Vector Machines, in terms of raw classification. Particularly, Random Forest performs better in terms of AUC and is perceived as well-balanced for parallel variations of sensitivity and specificity, particularly with table-shaped data based on sensor measurements. This makes it simpler to use the model when the amounts of data are limited and the categories are easy to interpret.

From this, an essential choice between accuracy and speed becomes apparent in making AI models for PdM. Although adaptable and presenting powerful visuals, deep learning models are commonly challenging to understand and require a substantial amount of data and powerful processing to achieve their full potential. However, classical ML models provide more precise explanations and are simpler to set up for calibration between precision and recall in live operations. For this reason, while CNNs outperform other models in many ways, classical ML remains competitive when working with small and well-organized datasets. Model calibration plays a critical role in this area. Models with high accuracy may still perform poorly if the scores indicating their confidence are not accurate. The results show that the Random Forest model outputs were more accurately calibrated than those from the augmented CNN model. Still, MC Dropout plays a vital role in this work because it supplies probabilistic information to overconfident black-box models, helping to improve decision-making in key maintenance areas.

6. Conclusion

A novel strategy for predictive maintenance is created by merging computer vision and probabilistic machine learning into a single process.

Using sensor time series, the pipeline generates synthetic images, on which CNNs can classify the states of a machine. To overcome the need for specific predictions, MC Dropout is joined with the CNN, making it possible to measure the model's faithfulness in manufacturing areas that focus on safety. For comparison, the system is evaluated against existing interpretable models, such as Random Forest and SVM, allowing us to examine their performance, calibration, and the level of confidence we have in their results. The primary objective of this study is that using accuracy alone is insufficient to evaluate the suitability of a model in Predictive Maintenance (PdM). When it comes to machine safety and extended equipment life, considering interpretability and calibration is crucial. Although CNN models allow for different levels of complexity, adding mechanisms such as MC Dropout is necessary to make their outcomes truly actionable. Future work will involve utilizing absolute sensor data from industrial cameras to enhance picture quality. Additionally, models that utilize CNNs to extract features and Random Forests as classifiers can be created. Furthermore, probabilistic RNNs will be examined to ensure time-based uncertainties are captured in the PdM system.

References

[1] Aravind, R., C.V. Shah, and M.D. Surabhi, Machine learning applications in predictive maintenance for vehicles: case studies. International journal of engineering and computer science, 2022. **11**(11).

[2] Khan, A.I. and S. Al-Habsi, Machine learning in computer vision. Procedia Computer Science, 2020. **167**: p. 1444-1451.

[3] Seliya, N., A. Abdollah Zadeh, and T.M. Khoshgoftaar, A literature review on one-class classification and its potential applications in big data. Journal of Big Data, 2021. **8**: p. 1-31.

[4] Serradilla, O., et al., Deep learning models for predictive maintenance: a survey, comparison, challenges and prospects. Applied Intelligence, 2022. **52**(10): p. 10934-10964.

[5] Sharpe, C., et al., A comparative evaluation of supervised machine learning classification techniques for engineering design applications. Journal of Mechanical Design, 2019. **141**(12): p. 121404.

[6] Fernandes, M., J.M. Corchado, and G. Marreiros, Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: a systematic literature review. Applied Intelligence, 2022. **52**(12): p. 14246-14280.

[7] T, G., R. V, and U.K. N. L, Detecting Security Threats in Wireless Sensor Networks using Hybrid Network of CNNs and Long Short-Term Memory. International Journal of Intelligent Systems and Applications in Engineering, 2023. **12**(1s): p. 704 - 722.

[8] Malali, N., Using Machine Learning to Optimize Life Insurance Claim Triage Processes Via Anomaly Detection in Databricks: Prioritizing High-Risk Claims for Human Review. International Journal of Engineering Technology Research & Management (IJETRM), 2022. **6**(06).

[9] Rishabh Rajesh, S., Exploring the Use of Cloud-Based AI and ML for Real-Time Anomaly Detection and Predictive Maintenance in Industrial IoT Systems. International Journal of Intelligent Systems and Applications in Engineering, 2023. **11**(4): p. 925 – 937.

[10] Raschka, S., J. Patterson, and C. Nolet, Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence. Information, 2020. **11**(4): p. 193.

[11] Yuan, C.J., et al., Expert Analysis for Multi-criteria Human-in-the-Loop Input Selection for Predictive Maintenance Model, in Materials, Design and Manufacturing for Sustainable Environment: Select Proceedings of ICMDMSE 2022. 2022, Springer. p. 461-473.

[12] Staartjes, V.E. and J.M. Kernbach, Foundations of Machine Learning-Based Clinical Prediction Modeling: Part III—Model Evaluation and Other Points of Significance, in Machine Learning in Clinical Neuroscience: Foundations and Applications. 2021, Springer. p. 23-31.

[13] Aseeri, A.O., Uncertainty-aware deep learning-based cardiac arrhythmias classification model of electrocardiogram signals. Computers, 2021. **10**(6): p. 82.

[14] Lambert, B., et al., Trustworthy clinical AI solutions: a unified review of uncertainty quantification in deep learning models for medical image analysis. arXiv preprint arXiv:2210.03736, 2022.