Probabilistic Uncertainty-Aware Decision Fusion of Neural Network for Bearing Fault Diagnosis

Atabak Mostafavi[1,2], Mohammad Siami[3], Andreas Friedmann[1], Tomasz Barszcz[4], Radoslaw Zimroz[5]

[1] *Fraunhofer Institute for Structural Durability and System Reliability LBF, Darmstadt, Hessen, 64289, Germany*
atabak.mostafavi@lbf.fraunhofer.de
andreas.friedmann@lbf.fraunhofer.de

[2]*Technische Universität Darmstadt, Department of Mechanical Engineering, Darmstadt, Hessen, 64287, Germany*
atabak.mostafavi @stud.tu-darmstadt.de

[3]*AMC Vibro Sp. z o.o., Pilotow 2e, 31-462, Kraków, Poland*
msiami@amcvibro.com

[4]*Faculty of Mechanical Engineering and Robotics, AGH University, Al. Mickiewicza 30, 30-059, Kraków, Poland*
tbarszcz@agh.edu.pl

[5]*Faculty of Geoengineering, Mining and Geology, Wroclaw University of Science and Technology, Na Grobli 15, 50-421, Wroclaw, Poland*
radoslaw.zimroz@pwr.edu.pl

ABSTRACT

Reliability is a central aspect of machine learning applications, especially in fault diagnosis systems, where only an accurate and reliable diagnosis system is economically justifiable, considering that any false diagnosis would lead to an increase in maintenance costs or a reduction in system efficiency. Recent advances in machine learning (ML) techniques have encouraged condition monitoring researchers to focus their efforts on finding suitable ML-based solutions for system condition assessment. However, to address the reliability issue, it is crucial to consider a larger amount of data measured by heterogeneous sensors on the system together with non-sensor information. The trend of data fusion has already started in other areas of ML application, and many of today's state-of-the-art models benefit from various types of fusion techniques to improve their accuracy. However, traditional classifiers do not provide any information about the prediction uncertainty, and they tend to show falsely high confidence when encountering low-quality data or previously unseen classes. Fusion of different data sources without considering the epistemic or aleatory uncertainty can lead to a deterioration of the result. Bayesian frameworks have traditionally been used to quantify uncertainty of systems; however, only recent advances made it possible to successfully implement Bayesian ML models.

The research methodology was investigated using the MAFAULDA dataset generated by SpectraQuest's Machinery Fault Simulator. This simulator experimentally simulated various bearing conditions, including normal operation and inner and outer ring bearing failures, at variable speeds. The dataset consists of 1951 instances measured using two triaxial accelerometers, a microphone, and a tachometer.

Diagnosis has been done via two multi label 1D Convolutional Neural Networks - each for a selected sensor - and their prediction along with their associated uncertainty quantity has been fused utilizing Bayesian model averaging. The methodology is capable of fusion of various decisions made based on different data sources and generate a unified decision with associated confidence level. Fusion process is uncertainty aware and application of 1D networks reduce the amount of data needed.

1. INTRODUCTION

1.2. Motivation behind the study

While Condition Monitoring Systems (CMS) have been extensively researched in recent years, the issue of their reliability has often been overlooked. It's crucial to recognize that CM relies on a complex system comprising sensors, acquisition devices, data analysis techniques, and expertise. Having in mind that any system that can fail, would eventually fail highlights the importance of reliability studies on CMS. This gap in research domain has prompted the authors to investigate the reliability of CMS with the aim of increasing awareness and attracting other scientists' attention to the issue of the reliability of CMS.

The main goal of condition monitoring is to enhance the detection of failures compared to traditional methods like periodic maintenance in a cost-effective manner [1]. Therefore, addressing uncertainty in models does not imply admitting their malfunction; rather, a CMS with high uncertainty may prevent severe failures and associated costs and casualties that could be overlooked by competitor approaches. Acknowledging and addressing sources and levels of uncertainty in any diagnosis system is essential, as uncertainty is an inevitable aspect. Providing this critical information can help operators to make informed decisions and conduct thorough risk analyses.

1.1. Condition monitoring background

Condition monitoring (CM) serves as a vigilant process or a precision instrument focused on the early detection of machinery faults, failures, and wear, aiming to minimize downtimes and maintenance costs while maximizing production output. By detecting failures in their early stages, CM optimizes maintenance planning and action, thereby mitigating the risk of escalating damage and catastrophic failures. Moreover, it enhances comprehension of machinery behavior, consequently refining maintenance practices and operational efficiency [2].

CM techniques typically involve continuous measurement of machinery indicators or signals (online CM) or periodic assessments at predetermined intervals (offline CM) to detect abnormal deviations from baseline signals, distinguishing them from normal operational variations or detecting any fault signature [3].

Many examples for the development of CM can be found in literature: [4] designed and developed an integrated wireless vibration sensing tool to monitor milling equipment, employing Support Vector Machine (SVM) for analysis. [5] compared the statistical parameters of vibration signals for bearing diagnosis and suggested that signal power is the most effective criterion for diagnosis. [6] proposed an intelligent feature extraction method from vibration signals of bearing datasets to prevent human intervention for large signal analysis tasks. [7] reviewed vibration based condition monitoring of rotary machinery. [8] proposed a deep learning based gearbox fault diagnosis method that addresses data scarcity. [9] have fused multiple vibration signal into two-dimensional rectangular matrix and employed a two-dimensional convolutional neural network (2D-CNN) for bearing fault diagnosis.

Rotating machinery is a primary focus of CM research due to its challenging nature. This includes various industrial components such as rolling and journal bearings, gearboxes, shafts, blades, entire systems like wind turbines, reciprocating machines, electric motors, pumps, helicopters, fans, cam mechanisms, generators, and compressors. Various diagnostic parameters can be monitored, such as vibrations, acoustic emissions, electrical currents, flow rates, rotational speeds, pressure levels, temperature, lubrication conditions, strain, wear, and rotor-stator interactions. Vibration emerges as the predominant condition indicative of rotary machine health, as each component exhibits a unique vibration signature closely correlated with operational conditions. Faults or defects within components introduce additional dynamic forces, manifesting as vibrations within specific frequency ranges. Notable fault types detectable via vibration-based CM techniques include looseness, eccentricity, unbalance, blade defects, misalignment, bearing faults, gear damage, and shaft deformations [3].

1.2. Uncertainty in diagnosis

Uncertainty plays a significant role in human affairs, permeating everyday decisions in ordinary life. Decision-making, a fundamental capability of human beings, is essential for survival and well-being. However, decision-making is inherently challenged by uncertainty about the future. Anticipation of future events, upon which decisions are based, is inevitably subject to uncertainty. This is particularly evident in diagnostic uncertainty in engineering, where engineers often struggle to make definitive diagnoses despite extensive testing and relevant information [10].

In the realm of CM, ensuring the reliability of the diagnostic system is paramount. Indicating a fault where no fault is (a so-called false positive) can lead to unnecessary stoppages and maintenance, increasing operational costs, while false negatives risk failure and the propagation of damage. Ensuring the reliability of CMS is crucial for achieving their main goals of cost reduction and failure prevention. Considering the substantial investment necessary for implementing these techniques, only a reliable system that effectively prevents expensive failures can be justified.

To address these challenges, an uncertainty-aware fusion approach is essential. This approach involves explicitly modeling and quantifying the uncertainty associated with each source of information and the fusion process itself. By accounting for uncertainty, decision-makers can better assess the reliability and confidence level of the fused information. Moreover, an uncertainty-aware fusion approach enables the identification of potential sources of error or bias in the fusion process, allowing for more robust and trustworthy decision-making outcomes.

Information fusion, as a methodological approach, presents a promising solution to the challenge of managing uncertainty in complex systems. By integrating diverse sources of information, including both sensory and non-sensory data, information fusion aims to enhance decision-making processes by providing a more comprehensive and accurate understanding of the observed system [11].

One of the primary advantages of information fusion is its capacity to leverage the strengths of individual sources of information while compensating for their inherent limitations. For instance, while sensory data such as vibration measurements may provide insights into the mechanical condition of a machine, non-sensory data such as operational logs or historical maintenance records can offer valuable contextual information. By combining different types of information, information fusion enables a more holistic assessment of the system's health status. However, implementing data fusion poses several challenges, particularly due to the diversity of data sources and sensor technologies involved. These challenges include issues related to data compatibility, data quality, and data integration. For instance, data collected from different sensors may vary in terms of accuracy, precision, and sampling frequency, making it challenging to effectively merge them into a cohesive dataset. Neglecting model uncertainty during fusion process can significantly impact the reliability of the fused information. Inaccurate or unreliable diagnoses from individual sources can propagate errors and inconsistencies throughout the fusion process, leading to a loss of fidelity in the final fused output. [12]

1.3. Authors contribution

The field of condition monitoring is vast, with numerous research initiatives aiming to enhance fault diagnosis techniques. This work contributes to the existing body of knowledge by introducing several approaches:

1- Multi-Label Fault Diagnosis: The authors propose a multi-label approach to fault diagnosis, enabling handling of complex fault scenarios. This methodology allows for the assignment of independent probability values to each fault class, providing a more detailed understanding of the system's health status.
2- Addressing Data Scarcity: The research addresses the common challenge of data scarcity by introducing a Custom 1D Convolutional Neural Network (CNN). 1D CNN architecture reduces the amount of data required for accurate fault diagnosis, thereby overcoming limitations associated with insufficient data availability.
3- Reliability Enhancement: The study enhances the reliability of fault diagnosis by leveraging multiple probabilistic decisions from different sensors. Through a Bayesian Model Averaging (BMA) approach, the authors combine the probabilistic outputs of various sensors, resulting in more robust and accurate diagnostic outcomes. This integration of diverse sensor data contributes to improved decision-making and system health assessment.

2. MULTILABEL PROBLEM

In traditional single-label classification, the model learns from a set of examples, each associated with a single label from a set of distinct labels. Typically, a traditional classifier utilizing a SoftMax layer assigns a probability value to each label, ensuring that the sum of probabilities across all possible labels equals one. The model then selects the label with the highest probability as the predicted label. However, this approach limits the model to predicting only one label per instance.

In contrast, in multi-label classification, the model assigns a probability value between zero and one independently to each class for a given instance. This allows for multiple labels to simultaneously have high probabilities. These classes are non-mutually exclusive and may overlap by definition. This approach was mainly used for text categorization and medical diagnosis [13]. Multi-label classification has been used by [14] to classify X-ray image via residual attention learning to diagnosis thorax disease. [15] utilized multi-label modeling for person re-identification to address the challenges of unsupervised learning, utilizing memory-based non-parametric classifier and integrates multi-label classification and single-label classification in a unified framework. [16] used attention based multi-label graph neural network to highlight the dependencies of labels in text classification.

In the context of system diagnosis and machinery fault detection, multi-label classification has not been widely investigated despite its value for identifying complex faults, especially when there's a correlation between them. By setting an appropriate threshold, the model can predict a neutral class, indicating uncertainty about the outcome, rather than forcing a specific label prediction. In contrast to single-label classification where the model assigns a probability value summing up to one across all classes, multi-label classification assigns an independent probability to each label (see Figure 1). This means that there may be cases where none of the labels have a high enough probability to cross the threshold, indicating that the network lacks confidence in its prediction.

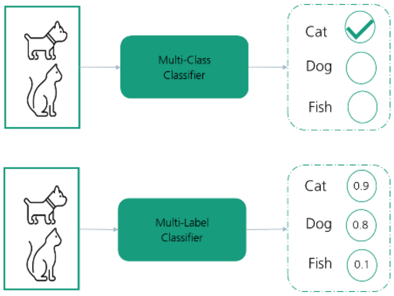

Figure 1 Label assignment in Multi-class vs Multi-label modeling

Various techniques exist for implementing a multi-label model, as shown in Figure 2. However, the details of each technique are beyond the scope of this paper, and interested readers are referred to [13, 17] for more information.

In addressing the problem, two primary approaches are commonly employed. Firstly, we can transform the problem into smaller, single-label components, allowing the use of any machine learning method to address each segment. Alternatively, we can adapt the algorithm itself to enable multi-label classification.

In the transformation approach, we can convert the problem into single-label binary classification using various methods:

- Powerset of Labels: This method decomposes the problem into all possible combinations of labels. While it provides insight into label relations, it can be computationally expensive.
- Binary Relevance: This approach compares a single label to all others or to one other label.
- Label Manipulation: We can also delete or create new labels as needed.

When implementing CNNs, different loss functions and activation layers may be required at the end of the network to accommodate multi-label classification.

3. BAYESIAN MODEL AVERAGING

In many cases, multiple models can adequately describe the distributions that generate observed data. When faced with this scenario, selecting the best model becomes crucial and is typically based on criteria such as how well the model fits the observed dataset, its predictive capabilities, or likelihood penalizations like information criteria. Once a model is selected, inferences are drawn and conclusions are made under the assumption that the selected model accurately represents the underlying truth. However, there are drawbacks to this approach. Selecting a single model can lead to overconfident inferences and riskier decisions, as it overlooks the inherent uncertainty in model selection and relies heavily on specific assumptions about the selected model. [18]

BMA provides a systematic and coherent methodology for addressing model uncertainty. It applies Bayesian inference directly to the problem of model selection, combined estimation, and prediction. BMA provides a straightforward criterion for model selection and leads to more cautious predictions. However, implementing BMA can be challenging, as it involves making various assumptions and decisions based on specific situations and contexts. [18]

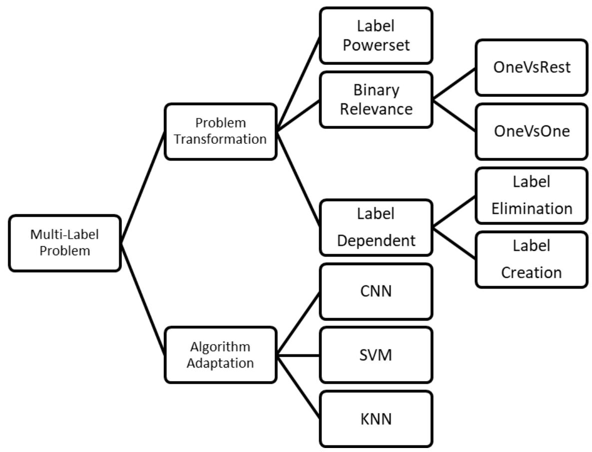

Figure 2 Overview of multi-label classification techniques

Let us consider an ensemble of models represented as $\boldsymbol{M_l}, \boldsymbol{l} = \mathbf{1}, \dots, \boldsymbol{K}$, and let $\boldsymbol{Y}$ represent observed data from dataset and $\boldsymbol{\theta_l}$ be parameter of the model $\boldsymbol{l}$, then likelihood function of $\boldsymbol{Y}$ given $\boldsymbol{\theta_l}$ and $\boldsymbol{M_l}$ can be written as $\boldsymbol{L(Y|\theta_l, M_l)}$. Additionally, prior probability of model parameters neglecting hyperparameters can be written as $\boldsymbol{\pi(\theta_l|M_l)}$ now, one can easily show posterior probability for model parameters as:

$$\pi(\theta_l|Y, M_l) = \frac{L(Y|\theta_l, M_l)\pi(\theta_l|M_l)}{\int L(Y|\theta_l, M_l)\pi(\theta_l|M_l)\, d\theta_l} \tag{1}$$

The denominator of (1) is called model's marginal likelihood or model evidence which represent prior distribution of all the parameter values related to model $\boldsymbol{M_l}$. Let's denote it as:

$$\pi(Y|M_l) = \int L(Y|\theta_l, M_l)\pi(\theta_l|M_l)\, d\theta_l \tag{2}$$

Bayesian model averaging introduces an additional level to this hierarchical modeling framework by incorporating a prior distribution over the entire set of models under consideration. This incorporates for the prior uncertainty regarding each model's ability to accurately represent the observed data. This is represented as a probability density function across all the models, and can be written as $\boldsymbol{\pi(M_l)}$

or $\boldsymbol{l} = \mathbf{1}, \dots, \boldsymbol{K}$, now we can show the posterior of model probability as:

$$\pi(M_l|Y) = \frac{\pi(Y|M_l)\pi(M_l)}{\Sigma_{l=1}^{k}\pi(Y|M_l)\pi(M_l)} \tag{3}$$

One now may re-write (3) as a ratio to a baseline model:

$$BF_{lm} = \frac{\pi(M_l|Y)}{\pi(M_m|Y)} \tag{4}$$

This can be interpreted as the relative strength of the models with respect to each other. It is clear that Eq. (3) can be expressed as the division of Eq. (4) as: [18]

$$\pi(M_l|Y) = \frac{BF_{l1}\pi(M_l)}{\Sigma_{m=1}^{k}BF_{m1}\pi(M_m)} \tag{5}$$

If $\boldsymbol{\Delta}$ is a quantity of interest, such as the utility of a course of action, then its posterior distribution can be formulated as: [19]

$$\pi(\Delta|Y) = \Sigma_{l=1}^{k}\pi(\Delta|M_l, Y)\pi(M_l|Y) \tag{6}$$

Here and on for simplicity we would address $\boldsymbol{\pi(M_l|Y)}$ term as $\boldsymbol{w_l}$. The $\boldsymbol{w_l}$s are probabilities; hence, they are nonnegative and sum up to 1. It is important to bear this in mind during their estimation. [20]

4. Estimating by likelihood maximization

For convenience, we restrict attention to the situation where the conditional probability density functions (PDFs) are approximated by normal distributions. We maximize $\boldsymbol{w_k}$ by maximum likelihood from the validation/training dataset. The likelihood function is defined as the probability of the training data given the parameters to be estimated. The maximum likelihood estimator is the value of the parameter vector that maximizes the likelihood function, that is, the value of the parameter vector under which the observed data were most likely to have been observed. It is convenient to maximize the logarithm of the likelihood function (or log-likelihood function) rather than the likelihood function itself, for reasons of both algebraic simplicity and numerical stability; the same parameter value that maximizes one also maximizes the other. Estimation through likelihood maximization involves approximating the conditional PDFs here for ease of computation normal distributions has been selected. We maximize the weights $\boldsymbol{w_k}$by maximizing the likelihood function using the validation/training dataset. The likelihood function represents the probability of observing the training data given the parameters to be estimated.

$$L(w_k|Y) = \Sigma_t \Sigma_{k=1}^{k} \log \pi(\Delta|M_l, Y)\, w_k \tag{7}$$

where the summation is over values of $\boldsymbol{t}$ that index observations in the training set. [20]

5. Model architecture

5.1. Convolutional neural network

CNNs have received considerable attention and have been proven effective in various domains. One promising area for CNNs is in fault diagnosis and CM. Researchers have been increasingly using ML techniques, especially CNNs, for system diagnosis, particularly when monitoring signals such as vibration, acoustics, or temperature. [21] has utilized multi-branch residual convolutional neural network to diagnose crane gearbox with vibration signal that has been transferred to 2D images using Markov transformation field. [22] suggested an explainable CNN model that analysis cyclostationary vibration signals to diagnose wind turbine gearbox fault. [23] has proposed a light weight CNN model for bearing fault diagnosis based on Fast Fourier Transfer

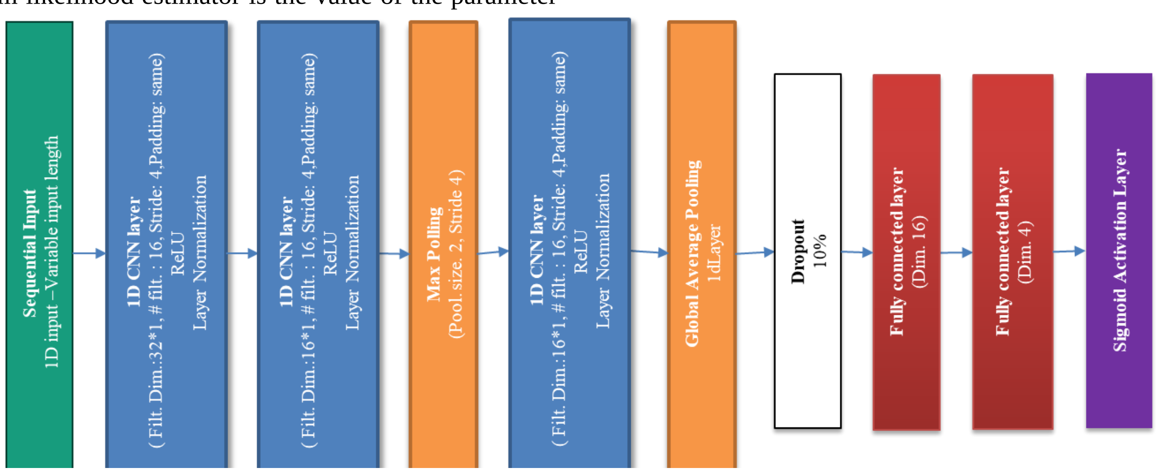

Figure 3 1D CNN for multi-label classification of bearing fault

(FFT) image coding of vibration signals. [24] has proposed a multiscale quadratic attention-embedded CNN with attention mechanisms to address the challenges associated with bearing vibration signals for fault diagnosis. [25] has fused vibration and microphone signals utilizing a 1D-CNN to enhance the accuracy of diagnosis. [26] has introduced a CNN model to diagnose bearing fault utilizing motor speed signal to remove the necessity of additional sensors.

A CNN consists of several layers, including an input layer, a convolutional layer, an activation layer, and a fully connected layer. Additional layers such as normalization and dropout are often used for generalization and to prevent overfitting. At the core of CNNs are convolutional layers, which allow us to automatically extract features from input data by mimicking how the brain's visual cortex processes images. This can be achieved by convoluting the input data with a filter, which is an n by m matrix whose elements are defined during the training phase, and moving the filter through the data at a constant step called a "stride". The convolution layer produces new images called feature maps. The feature map emphasizes the unique features of the original image. [27, 28] Although 2D CNNs have been commonly used for vibrational based diagnosis tasks, their effectiveness depends on a preprocessing step that converts the 1D signal into a 2D format. However, this preprocessing step often results in information loss and reduced diagnostic reliability. Although 1D and 2D CNNs share similar architectures, the key difference between them lies in their filter sliding mechanisms. In 1D CNNs, the filter slides vertically along the height to extract features, with the height determining the number of sample points for convolutional operations. On the other hand, 2D CNNs slide the filter both horizontally and vertically, with the height and width of the filter dictating the range of convolution operations for each step. However, 1D CNNs offer advantages over their 2D counterparts when processing 1D signals. This preference stems from several factors: [29]

- Computational complexity of 1D and 2D convolution calculations differ due to the fact that 1D CNN operates with one dimension less, resulting in significantly lower computational costs under identical conditions (same configuration, network, and hyperparameters).
- Reduced computational complexity makes 1D CNN suitable for low-cost real-time applications on smaller devices.
- Processing signals in the time domain eliminates the need for an additional step to convert a one-dimensional signal to a two-dimensional signal. This avoids adding irrelevant data and preserves the information present in the original data.

Here, we introduce a customized 1D CNN network (refer to Figure 3) along with the associated hyperparameters (see Table 1) for multi-label classification of bearing fault diagnosis. The application of 1D CNN allows us to employ shallower networks and avoids the inclusion of irrelevant information that may result from the conversion of 1D to 2D data.

By employing a sigmoid activation function at the last layer of the CNN architecture, along with a binary entropy loss function, the conventional multi-class CNN classifier has been transformed into a multi-label classifier that operates independently within each class and predicts whether the instance belongs to that class or not, as in a "one against all" strategy. This approach eliminates the need to train multiple networks for each label, thus reducing the necessity of large data and computational effort.

Table 1 Model Hyperparameters

Hyperparameter	Value
Mini batch size	25
Max epoch	50
Network selection (Early stoppage)	Minimum validation loss
optimizer	Adam
Learning rate	0.001
Loss Function	Binary cross-entropy
Padding	“Same”
Software	MATLAB

6. TEST DATASET AND PREPRATION

The methodology has been applied on the MAFAULDA dataset. The dataset consists of 1951 multivariate time-series acquired by sensors on SpectraQuest's Machinery Fault Simulator (MFS) Alignment-Balance-Vibration (ABVT). It includes six different simulated states: normal function, imbalance fault, horizontal and vertical misalignment faults, and inner and outer bearing faults. This heterogeneous dataset involves measuring acoustic and vibration signals, providing comprehensive insights into machinery behavior and fault diagnosis. Each measurement lasts for 5 seconds, with 49 measurements for normal conditions, 197 for horizontal misalignment with angles of 0.5, 1.0, 1.5, and 2.0 degrees, 301 for vertical misalignment with angles of 0.51, 0.63, 1.27, 1.40, 1.78, and 1.90 degrees, and 333 for mass imbalance of 6, 10, 15, 20, 25, 30, and 35 grams. Bearing faults have been combined with 5, 6, 20, and 35 grams of mass imbalance to enhance the effect of the fault. The available experiment specification includes details of used equipment's, including the SpectraQuest Inc. Alignment/Balance Vibration Trainer (ABVT) Machinery Fault Simulator (MFS), Industrial IMI Sensors accelerometers, Monarch Instrument MT-190 analog tachometer, and Shure SM81 microphone. Data acquisition parameters such as sensitivity, frequency range, and measurement range are specified for each sensor. Sequences are categorized based on fault types, with details on the number of sequences per fault category, load values, and degrees of misalignment. The database is openly accessible online, with links provided at [30] for downloading the entire dataset or specific parts corresponding to different fault types. Figure 4 depicts the data preparation process for training the models. Raw vibration signals from the tangential direction of the overhang (sensor number four) and underhang (sensor number seven) accelerometers, each corresponding to a different model, are inputted along with the tachometer signal. These signals are then divided into five successive parts of one second each. The first three rotations of each one-second signal are then extracted, resulting in variable vector lengths. Following, random noise is added to reduce signal quality to signal-to-noise ratio (SNR) level of 10. The data set is then randomly divided into training (60 %), validation (20 %), and test (20 %) sets to facilitate model evaluation and validation. Additionally, reducing the data to three revolutions per second helps to evaluate the model under more realistic conditions where acquiring large datasets may not be feasible.

7. RESULT

The proposed 1D CNN was trained using the preprocessed training set (see Figure 4) of data from sensors four and seven separately. Probability acceptance threshold of 0.5 was set for each label output by the models. The models were then evaluated on the test dataset, and the performance results for the tangential overhang accelerometer signal are reported in Table 2, while those for the tangential underhang accelerometer are shown in Table 3. The corresponding confusion matrices are depicted in Figure 5 and Figure 6. Subsequently, BMA was performed on the two models, where BMA parameters were computed by maximizing the likelihood using the validation dataset. The related values are reported in Table 4. Finally, the results of the combined model via BMA, with the same probability acceptance threshold of 0.5, are shown in Table 5, along with its confusion matrix in Figure 7. Two instances from the test set have been selected and reported in Table 6 to demonstrate the step-by-step improvement of the results:

- In case A, the underhung model shows the highest probability for the outer race fault, which is an incorrect label. However, overhung and the combined model correctly identifies the fault as a cage fault.
- Case B reports an instance where the underhung model correctly identifies the label, but the overhung model fails to do so. Once again, the combined model correctly classifies the instance.

The results indicate an increase in performance of almost 5 % over the overhang accelerometer model and an increase of 8 % over the underhang accelerometer model. Considering the confusion matrix and accuracy of each class for each model, the calculated BMA parameters were as expected.

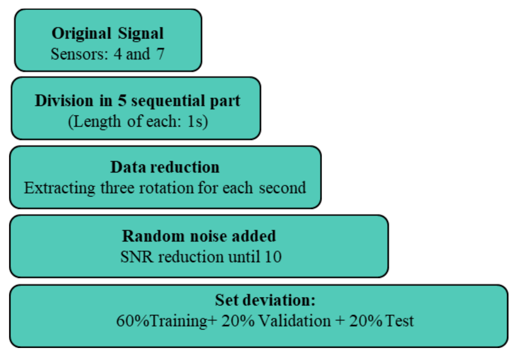

Figure 4 Data preparation scheme

Table 2 Accuracy of proposed 1D multi-label CNN for tangential Overhang accelerometer

Tangential Overhang accelerometer				
Label	Outer Race Fault	Cage Fault	Ball Fault	Healthy
Accuracy (%)	78.06	97.45	92.35	99.49
Overall Accuracy (%)			87.76	

Table 3 Accuracy of proposed 1D multi-label CNN for sensor tangential Underhung accelerometer

Tangential Underhung accelerometer				
Label	Outer Race Fault	Cage Fault	Ball Fault	Healthy
Accuracy (%)	95.41	89.29	95.41	99.49
Overall Accuracy (%)			84.69	

Table 4 BMA parameters

Posterior probability of Overhang Model	Posterior probability of Underhung Model
0.3728	0.6272

Table 5 Accuracy of combined model via BMA

Tangential Underhung accelerometer				
Label	Outer Race Fault	Cage Fault	Ball Fault	Healthy
Accuracy (%)	93.88	96.43	96.94	99.49
Overall Accuracy (%)			91.84	

Table 6 Instances from test set

	Case A			Case B		
	Overhang	Underhung	Combined	Overhang	Underhung	Combined
Outer Race Fault	0.04	0.59	0.39	0.46	0.78	0.66
Cage Fault	1.00	0.24	0.52	0.01	0.2	0.13
Ball Fault	0.00	0.05	0.03	0.54	0.02	0.21
Healthy	0.00	0.00	0.00	0.00	0.00	0.00
True label	Cage Fault			Outer Race Fault		

8. CONCLUSION

This study introduces a multi-label approach to fault diagnosis, which facilitate the handling of complex fault scenarios by assigning an independent probability value to each class. To address the common issue of data scarcity, a Custom 1D CNN is proposed to reduce the required amount of data. Additionally, a BMA approach is employed to enhance the reliability of diagnosis by combining multiple decisions from different sensors. Evaluation of the technique on a public dataset shows a 5 to 8 % improvement in the accuracy of the combined BMA model result compared to individual models. The discussed algorithm provides an explainable process for decision fusion, emphasizing the quality of each diagnosis. BMA offers an uncertainty-aware fusion platform, where each model contributes based on its performance in the training and validation phases.

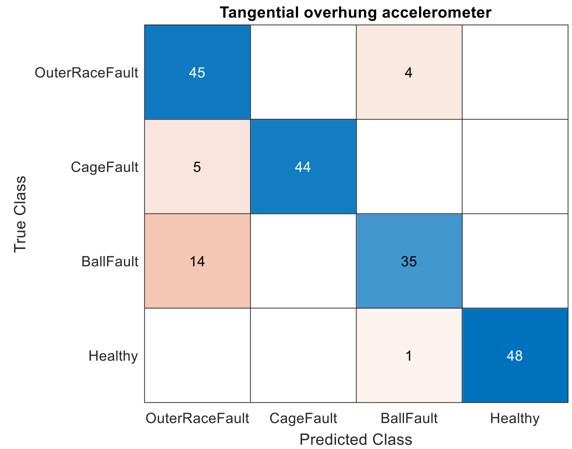

Figure 5 Confusion Matrix Multi-Label classifier - Tangential Overhung accelerometer

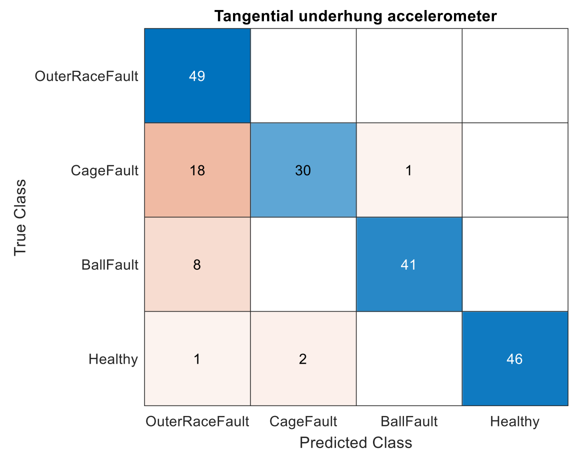

Figure 6 Confusion Matrix Multi-Label classifier - Tangential Underhung accelerometer

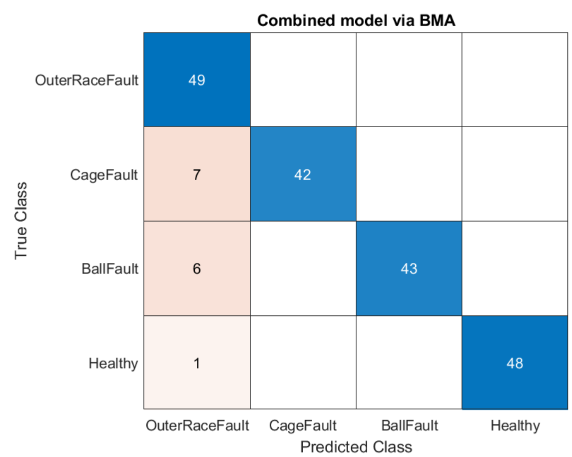

Figure 7 Confusion Matrix BMA combined Model

ACKNOWLEDGEMENT

The Authors Gratefully acknowledge the European commission for its support of the Marie Sklodowska Curie program through the ETN MOIRA project (GA 955681).

References

[1] David McMillan and Graham Ault, Eds., *Towards Quantification of Condition Monitoring Benefit for Wind Turbine Generators*, 2007.

[2] H. D. M. de Azevedo, A. M. Araújo, and N. Bouchonneau, "A review of wind turbine bearing condition monitoring: State of the art and challenges," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 368–379, 2016, doi: 10.1016/j.rser.2015.11.032.

[3] M. Tiboni, C. Remino, R. Bussola, and C. Amici, "A Review on Vibration-Based Condition Monitoring of Rotating Machinery," *Applied Sciences*, vol. 12, no. 3, p. 972, 2022, doi: 10.3390/app12030972.

[4] C. Zhou, K. Guo, and J. Sun, "An integrated wireless vibration sensing tool holder for milling tool condition monitoring with singularity analysis," *Measurement*, vol. 174, p. 109038, 2021, doi: 10.1016/j.measurement.2021.109038.

[5] N. Tandon, "A comparison of some vibration parameters for the condition monitoring of rolling element bearings," *Measurement*, vol. 12, no. 3, pp. 285–289, 1994, doi: 10.1016/0263-2241(94)90033-7.

[6] F. Jia *et al.,* "A method of automatic feature extraction from massive vibration signals of machines," in *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, 2016, pp. 1–6.

[7] C. Mongia, D. Goyal, and S. Sehgal, "Vibration response-based condition monitoring and fault diagnosis of rotary machinery," *Materials Today: Proceedings*, vol. 50, pp. 679–683, 2022, doi: 10.1016/j.matpr.2021.04.395.

[8] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, "Gearbox Fault Diagnosis Using a Deep Learning Model With Limited Data Sample," *IEEE Trans. Ind. Inf.*, vol. 16, no. 10, pp. 6263–6271, 2020, doi: 10.1109/TII.2020.2967822.

[9] J. Wang, D. Wang, S. Wang, W. Li, and K. Song, "Fault Diagnosis of Bearings Based on Multi-Sensor Information Fusion and 2D Convolutional Neural Network," *IEEE access : practical innovations, open solutions*, vol. 9, pp. 23717–23725, 2021, doi: 10.1109/ACCESS.2021.3056767.

[10] G. J. Klir, *Uncertainty and information: Foundations of generalized information theory*. Hoboken N.J.: Wiley-Interscience, 2006.

[11] S. Hassani, U. Dackermann, M. Mousavi, and J. Li, "A systematic review of data fusion techniques for optimized structural health monitoring," *Information Fusion*, vol. 103, p. 102136, 2024, doi: 10.1016/j.inffus.2023.102136.

[12] D. Wheeler, Esther D Meenken, Martin Espig, and Mos Sharifi, Eds., *UNCERTAINTY -WHAT IS IT?*, 2020.

[13] Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas, "A Review of Multi-Label Classification Methods,"

[14] Q. Guan and Y. Huang, "Multi-label chest X-ray image classification via category-wise residual attention learning," *Pattern Recognition Letters*, vol. 130, pp. 259–266, 2020, doi: 10.1016/j.patrec.2018.10.027.

[15] D. Wang and S. Zhang, "Unsupervised Person Re-identification via Multi-label Classification," 20-Apr-20. [Online]. Available: http:// arxiv.org/pdf/2004.09228

[16] A. Pal, M. Selvakumar, and M. Sankarasubbu, "MAGNET: Multi-Label Text Classification using Attention-based Graph Neural Network," *12th International Conference on Agents and Artificial Intelligence (ICAART*, pp. 494–505, 22-Mar-2020, doi: 10.5220/0008940304940505.

[17] A. C. P. L. F. de Carvalho and A. A. Freitas, "A Tutorial on Multi-label Classification Techniques," in *Studies in Computational Intelligence*, v. 201-206, *Foundations of computational intelligence*, A. E. Hassanien, Ed., Berlin: Springer, 2009-, pp. 177–195.

[18] T. M. Fragoso, W. Bertoli, and F. Louzada, "Bayesian Model Averaging: A Systematic Review and Conceptual Classification," *Int Statistical Rev*, vol. 86, no. 1, pp. 1–28, 2018, doi: 10.1111/insr.12243.

[19] Jennifer A. Hoeting, David Madigan, Adrian E. Raftery and Chris T. Volinsky, "Bayesian Model Averaging: A Tutorial,"

[20] ADRIAN E. RAFTERY, TILMANN GNEITING, FADOUA BALABDAOUI, and MICHAEL POLAKOWSKI, "Using Bayesian Model Averaging to Calibrate Forecast Ensembles,"

[21] J. Zhang, Q. Zhang, X. Qin, and Y. Sun, "Robust fault diagnosis of quayside container crane gearbox based on 2D image representation in frequency domain and CNN," *Structural Health Monitoring*, vol. 23, no. 1, pp. 324–342, 2024, doi: 10.1177/14759217231168877.

[22] A. Amin, A. Bibo, M. Panyam, and P. Tallapragada, "Wind Turbine Gearbox Fault Diagnosis Using Cyclostationary Analysis and Interpretable CNN," *J. Vib. Eng. Technol.*, vol. 12, no. 2, pp. 1695–1705, 2024, doi: 10.1007/s42417-023-00937-1.

[23] K. Cui, M. Liu, and Y. Meng, "A new fault diagnosis of rolling bearing on FFT image coding and L-CNN," *Meas. Sci. Technol.*, vol. 35, no. 7, p. 76108, 2024, doi: 10.1088/1361-6501/ad3295.

[24] Y. Tang, C. Zhang, J. Wu, Y. Xie, W. Shen, and J. Wu, "Deep Learning-Based Bearing Fault Diagnosis Using a Trusted Multiscale Quadratic Attention-Embedded Convolutional Neural Network," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–15, 2024, doi: 10.1109/TIM.2024.3374311.

[25] G. Li, J. Wu, C. Deng, Z. Chen, and X. Shao, "Convolutional Neural Network-Based Bayesian Gaussian Mixture for Intelligent Fault Diagnosis of Rotating Machinery," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021, doi: 10.1109/TIM.2021.3080402.

[26] Z. Guo, M. Yang, and X. Huang, "Bearing fault diagnosis based on speed signal and CNN model," *Energy Reports*, vol. 8, pp. 904–913, 2022, doi: 10.1016/j.egyr.2022.08.041.

[27] P. Kim, *MATLAB deep learning: With machine learning, neural networks and artificial intelligence*. New York, Berkeley, California: Springer; Apress, 2017. [Online]. Available: http://www.springer.com/de/book/9781484228456

[28] Neena Aloysius and Geetha M, *A Review on Deep Convolutional Neural Networks: 6th-8th April 2017, Melmaruvathur, India*. Piscataway, NJ, U.S.A.: IEEE, 2017. [Online]. Available: http://ieeexplore.ieee.org/servlet/opac?punumber=8269242

[29] R. F. R. Junior, I. A. d. S. Areias, M. M. Campos, C. E. Teixeira, L. E. B. Da Silva, and G. F. Gomes, "Fault detection and diagnosis in electric motors using 1d convolutional neural networks with multi-channel vibration signals," *Measurement*, vol. 190, p. 110759, 2022, doi: 10.1016/j.measurement.2022.110759.

[30] *MAFAULDA :: Machinery Fault Database [Online].* [Online]. Available: https://www02.smt.ufrj.br/~offshore/mfs/page_01.html (accessed: 06-Feb-24).