Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11 n.53, p. 82-89. May/June., 2025.
DOI: https://doi.org/10.5935/jetia. v11i53.1574

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

HEALTH CLASSIFICATION OF PUMPS USING TRANSFORMER-BASED DEEP LEARNING

Arunachalam Shivaa T V [1], Jayaprasanth D [2] and Arunshankar J [3]

[1,2,3] Department of Instrumentation and Control Systems Engineering, PSG College of Technology, Coimbatore, Tamil Nadu 641004, India.

1http://orcid.org/0009-0004-3385-2928, 2http://orcid.org/0009-0009-2189-4908, 3http://orcid.org/0000-0001-8244-0875

Email: arunachalam7shiva@gmail.com, djp.ice@psgtech.ac.in, jas.ice@psgtech.ac.in

ARTICLE INFO

Article History
Received: January 20, 2025
Revised: February 20, 2025
Accepted: May 15, 2025
Published: May 31, 2025

Keywords:
Pumps,
Health classification,
Sensor data,
LSTM,
Transformer.

ABSTRACT

This paper develops a health classification system for pumps to enhance operational efficiency and reduce unplanned downtime, crucial for manufacturing and water treatment industries. Leveraging real-time data from temperature sensors and industrial accelerometer, the system captures vital pump health indicators. Data is collected via Data Acquisition (DAQ) modules and by using Deep Learning (DL) techniques such as Long Short-Term Memory (LSTM) networks and Transformers; the pump health classification is achieved. These DL models excel at understanding complex temporal and spatial patterns in sensor data, essential for accurate fault detection. Through a comparative analysis of LSTM and Transformer models, their efficacy in pump health classification is assessed. This approach emphasizes the importance of sophisticated data analysis and deep learning in industrial maintenance practices. By providing fault detection, the system aims to significantly reduce maintenance costs, optimize resource usage, and enhance the safety and reliability of industrial operations.

I. INTRODUCTION

In the realm of industrial machinery and manufacturing, ensuring the seamless operation of critical components is of paramount significance. Industrial pumps, the workhorses in numerous sectors, including energy production, water treatment, and manufacturing processes, are indispensable for the continuous flow of materials. However, these mechanical workhorses are susceptible to wear, tear, and potential malfunctions. The application of data analytics has emerged as a transformative force in the sphere of predictive maintenance, offering the capacity to proactively assess and maintain the health of these pumps in real-time.

Pump failure detection is a critical task in many industrial applications, as it can help to prevent costly downtime and catastrophic failures. In recent years, deep learning methods have been shown to be effective in pump failure detection, achieving state-of-the-art results. Deep learning methods, particularly Long Short-Term Memory (LSTM) and Transformer models, have excelled in this task. LSTM is suitable for sequential data like sensor readings obtained from the pumps, but have limitations in capturing long-range dependencies. In contrast, Transformers, a newer neural network type, can process data in parallel, enhancing efficiency and the ability to model long dependencies.

This study encompasses the real-time data collected on-site, thereby presenting a comprehensive approach to formulating a dependable and efficient health classification system, for these indispensable components of industrial operations.

II. RELATED WORKS

The need for advanced fault detection and predictive maintenance in industrial systems has led to significant research into sensor technologies, Machine Learning (ML) and Deep Learning (DL). Pioneered this effort by utilizing smart sensors for monitoring centrifugal pumps, demonstrating the potential of real-time data in early fault detection [1]. This study set a precedent for integrating diverse sensor data with analytical models to enhance predictive maintenance strategies. Expanding on this foundation, the utility of vibration and motor current signature analysis (MCSA) in detecting faults in centrifugal pumps were carried out [2]. This work highlighted the complexity of interpreting the signals associated with mechanical failures, advocating for more nuanced diagnostic tools. A novel approach of using infrared thermography for predictive maintenance was introduced offering a non-invasive technique to monitor temperature variations indicative of underlying conditions, thereby broadening the spectrum of fault detection methods [3]. In the realm of ML, demonstrated the effectiveness of MLP and SVM algorithms in fault prediction within the oil and gas industry [4]. This study emphasized the critical role of algorithm selection in developing predictive models tailored to specific industrial contexts and data characteristics. A data-driven approach to predict pump failures, leveraging correlation analysis and empirical data was developed [5]. This methodology underscored the importance of integrating expert insights with analytical models to improve predictive accuracy.

The advancement of DL in fault diagnosis was significantly marked by Gamboa and utilized LSTM networks for time-series analysis. This approach addressed the challenges of analyzing temporal data, providing a robust framework for anomaly detection and forecasting [6]. Sabir *et al.* further validated the effectiveness of LSTM networks in diagnosing bearing faults in electrical machines, showcasing these models' capability to capture complex, time-dependent patterns characteristic of mechanical faults [7]. Diffusion-convolutional neural network (DCNN) was used for diagnosing pump faults from vibration data and high diagnostic accuracy was achieved in [8]. This work illustrated the potential of combining spatial analysis with traditional diagnostic data. The transformative potential of the Transformer model for fault diagnosis was introduced in [9]. The Anomaly Transformer model demonstrating the applicability of advanced DL models in industrial systems for fault identification was developed [10].

Further research by [11] and [12] expanded the application of ML and DL in industrial pump anomaly detection and bearing fault diagnosis, respectively. These studies highlighted the enhanced detection accuracy and efficiency afforded by advanced models. A comprehensive review of ML approaches for diagnosing faults in rotating equipment, emphasizing the superior performance of DL networks over traditional algorithms was carried out [13].

Sunal *et al*. reviewed ML-based fault detection for centrifugal pump induction motors, illustrating the ongoing advancements in the field and the importance of data quality and model selection [14]. The integration of CNNs with LSTMs marked a significant advancement in fault diagnosis, combining spatial and temporal data analysis for accurate fault identification [15],[16]. A transformer-based approach for novel fault detection was introduced showcasing the real-world applicability of advanced DL techniques in manufacturing and the effectiveness of these models in diverse applications [17]. Studies on centrifugal pump impeller crack detection, DL applications in rotating machinery fault diagnosis, and DL-based fault diagnosis of main pumps in converter stations were carried out [18-20].

The use of time series transformers for fault diagnosis in rotating machinery, demonstrating these models' ability to directly process time-series data and enhance fault identification accuracy was achieved [21],[22]. Reference [23] focused on fault classification of three-phase induction motors using Bi-LSTM networks, underscoring the potential of DL models in developing accurate and efficient fault classification systems. Markov parameters for fault detection in centrifugal pumps, presenting an innovative approach to fault diagnosis based on vibration data analysis was utilized in [24]. Applied a Transformer Neural Network for AC series arc-fault detection, illustrating the specific applications of DL models in addressing critical fault detection challenges in electrical systems [25].

Collectively, these studies form a comprehensive foundation for the development of advanced fault detection and predictive maintenance systems in industrial contexts. From the initial integration of sensor data with analytical models [1] to the application of advanced DL techniques for real-time fault identification [10],[17],[25], this field of research significantly advances the capabilities for fault detection, thereby enhancing operational efficiency, reducing maintenance costs, and improving safety across various industrial domains.

This paper is organized as follows: Section 3 gives description on the hardware setup. Proposed methodology is presented in Section 4. The performance of LSTM and Transformer models in classifying the health status of pump are discussed in Section 5. Finally, conclusions are drawn in Section 6.

III. HARDWARE DESCRIPTION

Figure 1 depicts the hardware setup developed for this work. The setup contains four similar centrifugal pumps. The leftmost pump is the healthy pump and the other three pumps are faulty. The second pump from the left has a bearing fault, and the third pump from the left has an impeller fault. The rightmost pump has both bearing and impeller faults. These faults are purposefully induced for destructive testing. Temperature and vibration sensors are mounted on these pumps to collect temperature and vibration data for fault detection. In this work, the leftmost pump (healthy) and rightmost pump (faulty) are considered for health classification problem using deep learning techniques.

Figure 1: Hardware setup.
Source: Authors, (2025).

The closer view of one of the pumps with sensors mounted is shown in Figure 2. An industrial accelerometer is mounted on top of the pump’s casing, while a J-type thermocouple is placed inside the pump's body and sealed in such a way that the thermocouple lead touches the bearings. The outputs of these sensors are connected to the respective signal conditioning modules, which are then acquired and stored in a computer running data logging software. The logged data is analysed subsequently. This strategic assembly of sensors and modules collectively captures a detailed picture of the pumps operational health, enabling a targeted and effective maintenance regime that bolsters industrial efficiency and reliability.

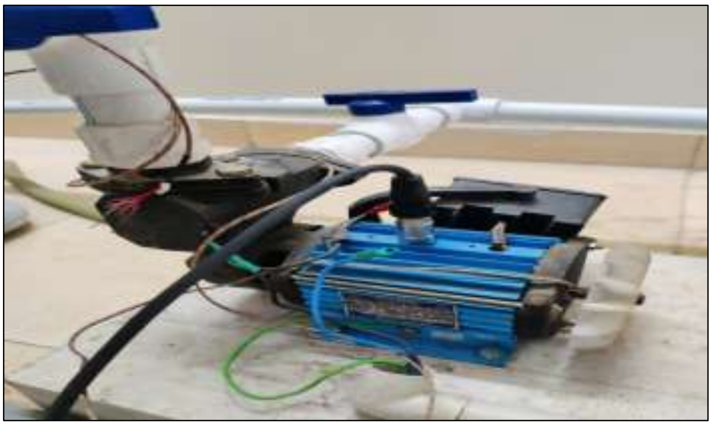

Figure 2: Closer view of a pump with sensors.
Source: Authors, (2025).

IV. PROPOSED METHODOLOGY

This section deals with the step-by-step implementation of the proposed method. It also briefly explains the DL algorithm implementation.

IV.1 FLOW DIAGRAM

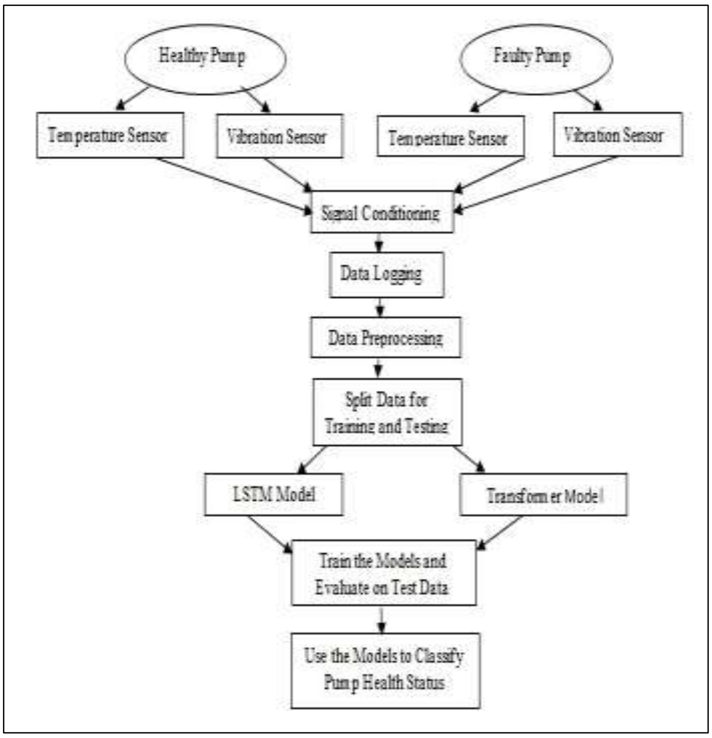

Figure 3: Work flow diagram.
Source: Authors, (2025).

Figure 3 shows the flow diagram that illustrates a process flow for a pump health monitoring system using deep learning models. The system differentiates between healthy and faulty pumps by analyzing the data collected from temperature and vibration sensors connected to the pumps. The process begins with signal conditioning, where the raw sensor data is filtered and transformed to a usable format using the DAQ modules. This is followed by data logging, where the conditioned sensor signals are stored and then data preprocessing is performed.

The preprocessed data is then divided into two sets: one for training the models and another for testing their performance. Two types of deep learning models are used: Long Short-Term Memory models and Transformer models. These models are trained separately with the training data set. After training, the models are evaluated using the test data to determine their accuracy and effectiveness in fault classification. Finally, the trained models are used to classify the health status of pumps in real-time or within a monitoring period, determining whether they are operating correctly (healthy) or have anomalies that suggest faults (faulty). The outcome of this process informs maintenance decisions, potentially leading to proactive interventions that can prevent breakdowns and maintain operational efficiency.

IV.2 ALGORITHM IMPLEMENTATION

Deep learning models like Long Short-Term Memory and Transformer have gained widespread popularity and proven to be highly effective for time series problems, surpassing traditional models for several key reasons.

IV.2.1 LONG SHORT-TERM MEMORY MODEL

Long Short-Term Memory is a type of recurrent neural network (RNN) architecture that has revolutionized the field of deep learning, particularly in the domain of sequential data processing. LSTMs address the vanishing gradient problem, which often hindered the training of traditional RNNs, by introducing a sophisticated memory mechanism that enables them to capture and remember long-range dependencies in sequences. LSTMs achieve their ability to learn long-term dependencies by using a combination of forget, input, and output gates that control the flow of information within the network. The forget gate determines what information from the previous state to forget, the input gate controls what new information to add, and the output gate controls what information to output. This allows LSTMs to selectively update, store, or discard information as they process data over time [15],[23].

The LSTM model design steps are as follows:

Step 1: Determine the number of LSTM layers: The number of LSTM layers depends on the complexity of the task and the size of the dataset. For small datasets, it is typically best to start with one or two LSTM layers. For larger datasets, or for more complex tasks, it may be necessary to use three or more LSTM layers.
Step 2: Determine the number of units in each LSTM layer: The number of units in each LSTM layer represents the complexity of the model. For small datasets, it is typically best to start with a small number of units. For larger datasets, or for more complex tasks, it may be necessary to use a larger number of units.
Step 3: Determine the activation function for the LSTM layers: The activation function for the LSTM layers controls the information flow through the network. The most common activation function for LSTM layers is the hyperbolic tangent (tanh) function. However, other activation functions, such as the sigmoid and rectified linear unit (ReLU) function, can also be used.
Step 4: Determine the optimizer for the LSTM model: The optimizer controls the parameters of the LSTM model which are updated during training. Some popular optimizers for LSTM models include Adam and RMSprop.
Step 5: Determine the loss function for the LSTM model: The loss function measures how well the LSTM model is performing

on the training data. Some common loss functions for LSTM classification tasks include cross-entropy and binary cross-entropy.

Step 6: Determine the batch size for the LSTM model: The batch size controls how many samples are processed by the LSTM model at each training step. A larger batch size can improve the efficiency of training, but it can also lead to overfitting. A smaller batch size can help to prevent overfitting, but it can also make training slower.

Step 7: Determine the number of epochs for the LSTM model: The number of epochs controls how many times the LSTM model is trained over the entire training dataset. A larger number of epochs can improve the performance of the model, but it can also lead to overfitting. A smaller number of epochs can help to prevent overfitting, but it may not allow the model to learn the training data fully.

IV.2.2 TRANSFORMER MODEL

The Transformer model represents a significant breakthrough in deep learning. Its fundamental innovation lies in the attention mechanism, enabling the simultaneous capture of long-range dependencies in input sequences through self-attention layers, resulting in more efficient computation, faster training, and improved accuracy. Furthermore, transformers have found utility in time series analysis, as demonstrated by models like Informer. Informer leverages self-attention to capture long-range temporal dependencies and incorporates positional encoding to discern the temporal order within time series data. This adaptability showcases the far-reaching impact of Transformer-based models in diverse fields, including time series forecasting [12],[25].

The transformer model design steps are as follows:

Step 1: Determine the number of encoder layers: The number of encoder layers depends on the complexity of the task and the size of the dataset.

Step 2: Determine the number of attention heads: The number of attention heads also depends on the complexity of the task and the size of the dataset.

Step 3: Determine the embedding dimension: The embedding dimension should be chosen to be large enough to capture the complexity of the input and output data.

IV.2.3 DESIGN PARAMETERS

Two deep learning models, LSTM and Transformer, were developed to classify the health status of industrial pumps based on separate temperature and vibration datasets. Table 1 and Table 2 represent the design parameters of the LSTM model and transformer model respectively.

The LSTM model is constructed with one layer containing 50 units, employing a sigmoid activation function, with Adam optimizer, and binary cross entropy as the loss function. It is trained over 5 epochs with a batch size of 128, processing 100,000 vibration data points and 23,701 temperature data points for both healthy and faulty pump conditions. The transformer model, designed for capturing complex dependencies, includes 2 encoder layers with 2 attention heads, and an embedding dimension of 128. It follows the same training regimen of 5 epochs and uses an identical dataset structure as the LSTM model. These design parameters enable the models to learn from the temporal and spatial patterns inherent in the sensor data, aiming to provide accurate fault classification to enhance industrial maintenance practices.

Table 1: Design parameters for LSTM model.

Design Parameters	Values/Function
No. of LSTM layers	1
No. of units in each layer	50
Activation function	Sigmoid
Optimizer	Adam
Loss function	Binary Cross Entropy
Batch Size	128
No. of epochs	5
Vibration data points	100000 (for each Healthy and Faulty Pump)
Temperature data points	23701 (for each Healthy and Faulty Pump)

Source: Authors, (2025).

Table 2: Design parameters for transformer model.

Design Parameters	Values/Function
No. of encoder layers	2
No. of attention heads	2
Embedding dimension (Head size)	128
No. of epochs	5
Vibration data points	100000 (for each Healthy and Faulty Pump)
Temperature data points	23701 (for each Healthy and Faulty Pump)

Source: Authors, (2025).

V. RESULTS AND DISCUSSIONS

V.1 LSTM MODEL FOR VIBRATION DATASET

The vibration data obtained from industrial accelerometer is used for training the model. Figure 4 represents the accuracy of the LSTM model for vibration data. It is inferred from Figure 4 that the initial high peak indicates that the model is learning well from the training data and performs effectively on both the training and validation datasets. The downturn in both training and validation accuracy implies a common issue affecting generalization which could be due to overfitting or a sudden change in the data distribution. The subsequent steady increase in accuracy for both datasets suggests that the model is recovering and adapting to the challenges presented during the downturn. As the number of epochs increases, the accuracy also increases.

A confusion matrix in binary classification provides a detailed breakdown of the model's performance by categorizing the predictions into four categories: True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN).

Figure 5 represents the confusion matrix of the LSTM model for the predictions made upon the test data. The TP and FP are 18,537 and 19,959 respectively. The TN and FN are 1,484 and 0 respectively. These values are used to find several other performance indicators of the model, such as Precision, Recall, and F1 Score.

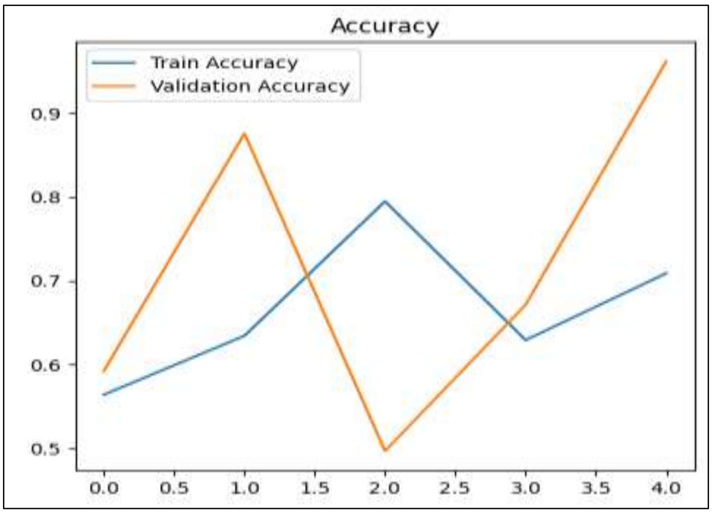

Figure 4: Accuracy of LSTM for vibration dataset.
Source: Authors, (2025).

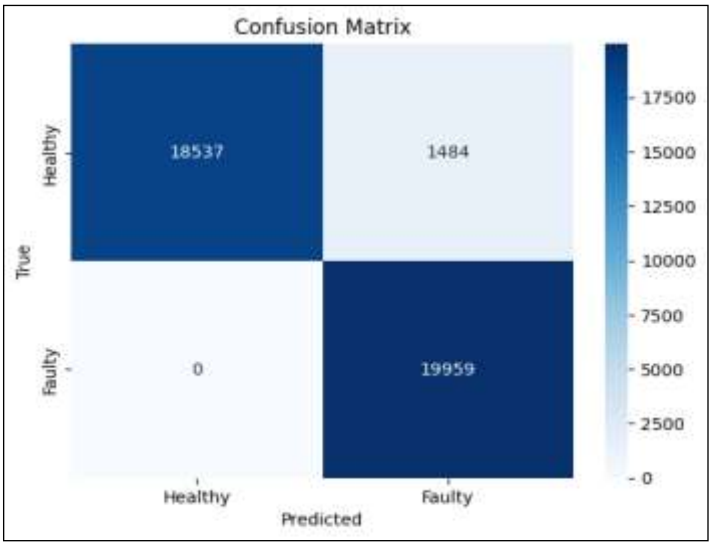

Figure 5: Confusion matrix of LSTM for vibration dataset.
Source: Authors, (2025).

Table 3 shows the classification report of the LSTM model indicating the key performance indicators. It also provides a comprehensive summary of various performance metrics such as precision, recall and F1-score for each class in a classification problem.

Table 3: Classification report of LSTM for vibration dataset.

Class	Precision	Recall	F1-score	Support
Faulty	1.00	0.93	0.96	20021
Healthy	0.93	1.00	0.96	19959
Accuracy			0.96	39980
Macro average	0.97	0.96	0.96	39980
Weighted average	0.97	0.96	0.96	39980

Source: Authors, (2025).

V.2 LSTM MODEL FOR TEMPERATURE DATASET

Figure 6 represents the accuracy of the LSTM model for temperature data. Figure 7 represents the confusion matrix of the predictions made upon the test data.

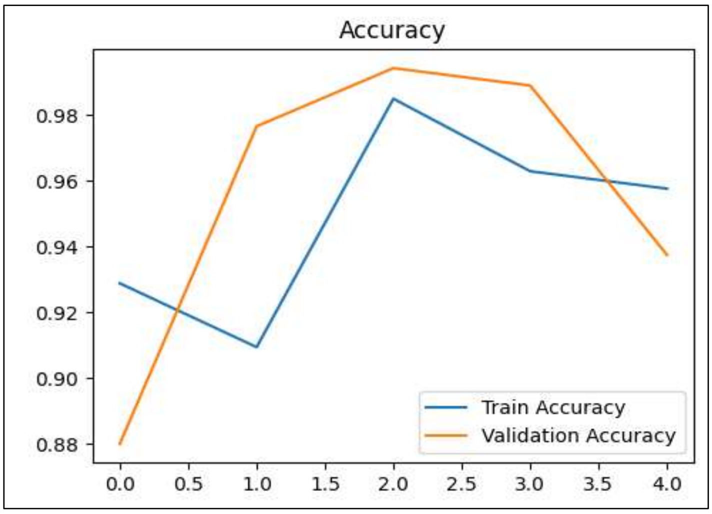

Figure 6: Accuracy of LSTM for temperature dataset.
Source: Authors, (2025).

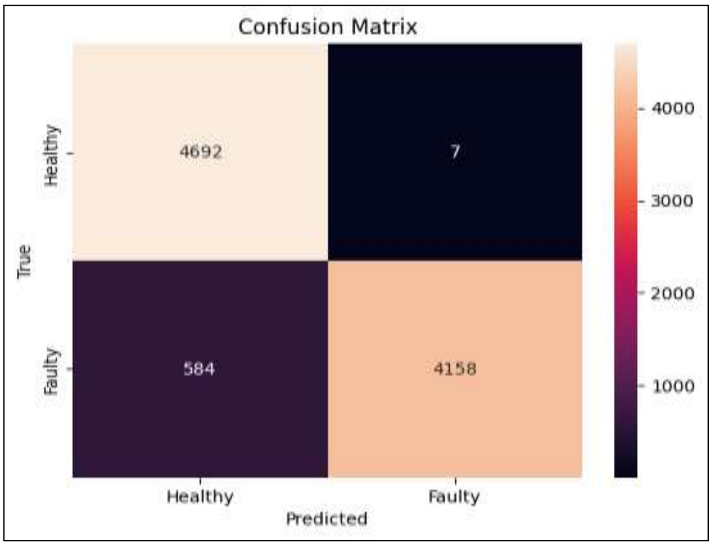

Figure 7: Confusion matrix of LSTM for temperature dataset.
Source: Authors, (2025).

Table 4 shows the classification report of the LSTM model indicating the key performance indicators. It is observed that the model made a good number of correct predictions of the pump status on the test data with an accuracy of 94%.

Table 4: Classification report of LSTM for temperature dataset.

Class	Precision	Recall	F1-score	Support
Faulty	0.89	1.00	0.94	4699
Healthy	1.00	0.88	0.93	4742
Accuracy			0.94	9441
Macro average	0.94	0.94	0.94	9441
Weighted average	0.94	0.94	0.94	9441

Source: Authors, (2025).

V.3 TRANSFORMER MODEL FOR VIBRATION DATASET

Figure 8 represents the accuracy of the transformer model for vibration data. It shows a higher level of accuracy in classifying the status of the pump. Figure 9 represents the confusion matrix of the predictions made upon the test data.

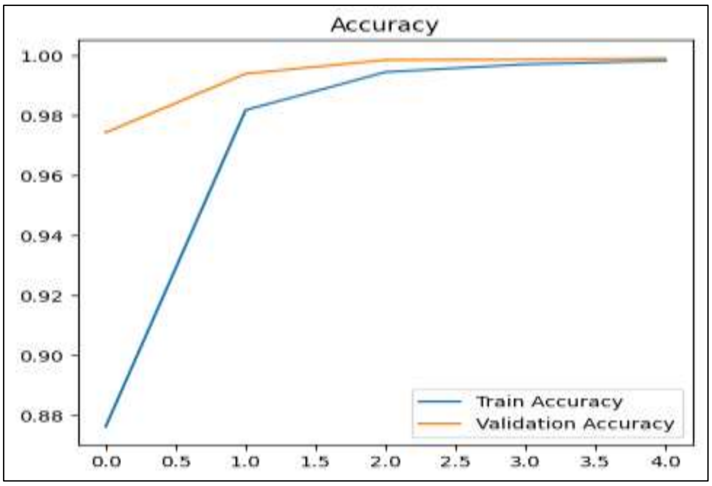

Figure 8: Accuracy of transformer model for vibration dataset.
Source: Authors, (2025).

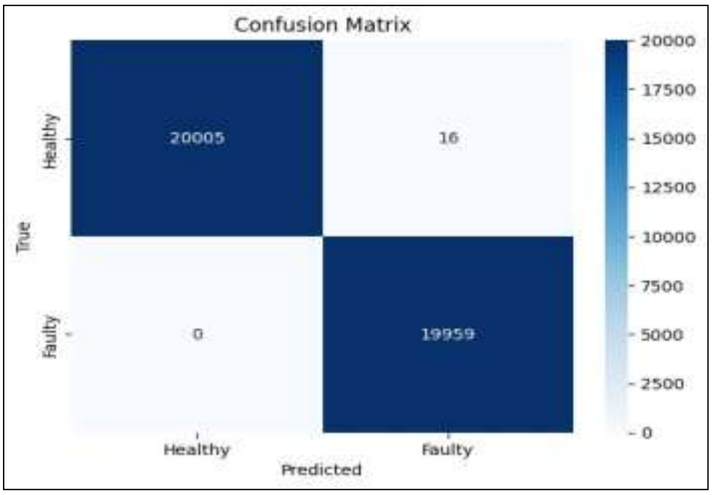

Figure 9: Confusion matrix of transformer model for vibration dataset.
Source: Authors, (2025).

Table 5 shows classification report of transformer model indicating key performance indicators. It is inferred that transformer model for vibration dataset achieved an extremely high accuracy of around 100% in classifying the pump health status.

Table 5: Classification report of transformer model for vibration dataset.

Class	Precision	Recall	F1-score	Support
Faulty	1.00	1.00	1.00	20021
Healthy	1.00	1.00	1.00	19959
Accuracy			1.00	39980
Macro average	1.00	1.00	1.00	39980
Weighted average	1.00	1.00	1.00	39980

Source: Authors, (2025).

V.4 TRANSFORMER MODEL FOR TEMPERATURE DATASET

Figure 10 represents the accuracy of transformer model on temperature data. It shows that greater accuracy is achieved in classifying the status of the pump compared to LSTM model. Figure 11 represents confusion matrix of the predictions made upon the test data.

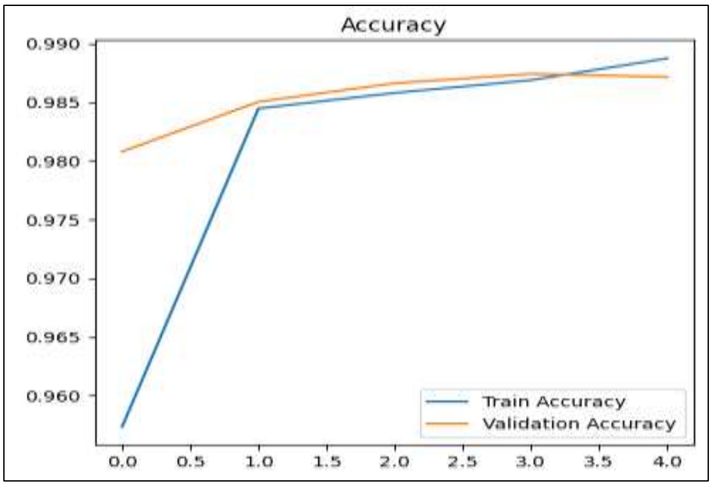

Figure 10: Accuracy of transformer model for temperature dataset.
Source: Authors, (2025).

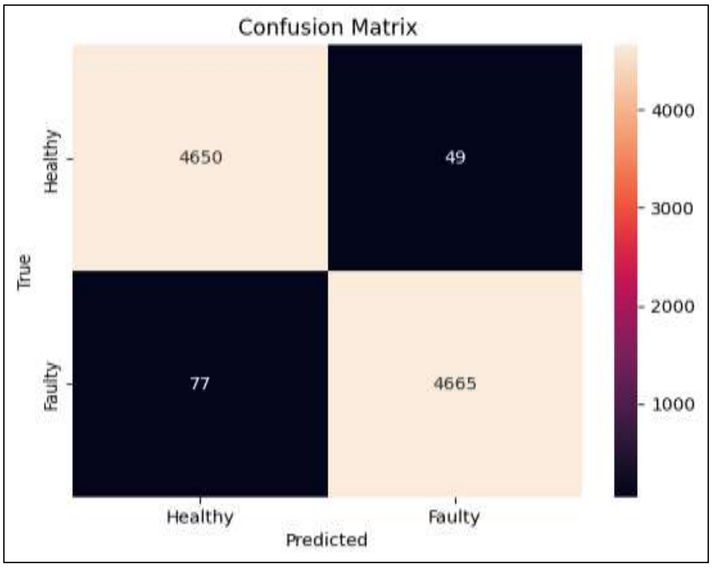

Figure 11: Confusion matrix of transformer model for temperature dataset.
Source: Authors, (2025).

Table 6 shows the classification report of the transformer model indicating the key performance indicators. The transformer model for temperature dataset achieved a very good accuracy of 99% in classifying the pump health status.

Table 6: Classification report of transformer model for temperature dataset.

Class	Precision	Recall	F1-score	Support
Faulty	0.99	0.99	0.99	4699
Healthy	0.99	0.99	0.99	4742
Accuracy			0.99	9441
Macro average	0.99	0.99	0.99	9441
Weighted average	0.99	0.99	0.99	9441

Source: Authors, (2025).

V.5 COMPARATIVE ANALYSIS OF LSTM AND TRANSFORMER MODELS

It is inferred that from Table 7 and Table 8, the classification accuracy obtained with the Transformer model is comparatively better than LSTM model. Hence, the Transformer model outperforms the LSTM model for pump health classification.

Table 7: Comparison of LSTM and Transformer model for vibration dataset.

Performance Metrics	LSTM	Transformer
Accuracy	0.9629	0.9996
Precision	0.9308	0.9992
Recall	0.9995	1.0000
F1-score	0.9642	0.9996

Source: Authors, (2025).

Table 8: Comparison of LSTM and Transformer model for temperature dataset.

Performance Metrics	LSTM	Transformer
Accuracy	0.9374	0.9867
Precision	0.9983	0.9896
Recall	0.8768	0.9858
F1-score	0.9336	0.9867

Source: Authors, (2025).

V. CONCLUSIONS

The comparative performance analysis of the LSTM and Transformer models for pump health classification using vibration and temperature datasets reveals a distinct advantage in favor of the Transformer model. These findings underscore the effectiveness of the Transformer architecture in handling sequential data, benefitting from its attention mechanisms that capture long-range dependencies more effectively than LSTM's gated recurrent units. Consequently, the Transformer model proves to be a robust and highly accurate tool for predictive maintenance in industries, offering significant potential to reduce downtime and maintenance costs through timely and precise fault detection.

In conclusion, the empirical evidence from this study strongly supports the Transformer-based models over LSTMs for monitoring the health of industrial pumps using sensor data. The higher accuracy and precision of the Transformer model can enable more reliable and effective predictive maintenance strategies, contributing to the advancement of smart monitoring practices.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Arunachalam Shivaa T V , Jayaprasanth D and Arunshankar J
Methodology: Arunachalam Shivaa T V , Jayaprasanth D and Arunshankar J
Investigation: Arunachalam Shivaa T V , Jayaprasanth D and Arunshankar J
Discussion of results: Arunachalam Shivaa T V , Jayaprasanth D and Arunshankar J
Writing Original Draft: Arunachalam Shivaa T V , Jayaprasanth D and Arunshankar J
Writing Review and Editing: Arunachalam Shivaa T V , Jayaprasanth D and Arunshankar J
Supervision: Arunachalam Shivaa T V , Jayaprasanth D and Arunshankar J
Approval of the final text: Arunachalam Shivaa T V , Jayaprasanth D and Arunshankar J

VII. ACKNOWLEDGMENTS

This work is funded by All India Council for Technical Education (AICTE) - Research Promotion Scheme (RPS) under File No. 8 – 73/ FDC/ RPS (POLICY – 1)/ 2019 – 20.

VIII. REFERENCES

[1] Lei Chen, Lijun Wei, Yu Wang, Junshuo Wang, and Wenlong Li, "Monitoring and predictive maintenance of centrifugal pumps based on smart sensors", *Sensors*, vol. 22, no. 6, p. 2106, 2022. https://doi.org/10.3390/s22062106

[2] A. R. Mohanty, P. K.Pradhan, N. P. Mahalik, and S. G. Dastidar, "Fault detection in a centrifugal pump using vibration and motor current signature analysis", *Int. J. Autom. Control*, vol. 6, no. 3–4, pp. 261–276, 2012. https://doi.org/10.1504/IJAAC.2012.051884

[3] A. D. Marinescu, C. Cristescu, T. C. Popescu, and C. A. Safta, "Assessing the opportunity to use the infrared thermography method for predictive maintenance of hydrostatic pumps", in *Proc. Int. Conf. on Energy and Environment (CIEM)*, 2017, pp. 270–274. https://doi.org/10.1109/CIEM.2017.8120790

[4] Orrù, Pier Francesco, Andrea Zoccheddu, Lorenzo Sassu, Carmine Mattia, Riccardo Cozza, and Simone Arena, "Machine learning approach using MLP and SVM algorithms for the fault prediction of a centrifugal pump in the oil and gas industry", *Sustainability*, vol. 12, no. 11, p. 4776, 2020. https://doi.org/10.3390/su12114776

[5] A. Mohammed, "Data driven-based model for predicting pump failures in the oil and gas industry", *Eng. Fail. Anal.*, vol. 145, p.107019, 2023. https://doi.org/10.1016/j.engfailanal.2022.107019

[6] J. C. B. Gamboa, "Deep learning for time-series analysis", arXiv preprint arXiv:1701.01887, 2017. https://doi.org/10.48550/arXiv.1701.01887.

[7] R. Sabir, D. Rosato, S. Hartmann, and C. Guehmann, " LSTM based bearing fault diagnosis of electrical machines using motor current signal", in *Proc. Int. Conf. on Machine Learning and Applications (ICMLA)*, 2019. https://doi.org/10.1109/ICMLA.2019.00113

[8] S. Manikandan, and K. Duraivelu, "Vibration-based fault diagnosis of broken impeller and mechanical seal failure in industrial mono-block centrifugal pumps using deep convolutional neural network", *J. Vib. Eng. Technol.*,vol. 11, no.1, pp. 141-152, 2023. https://doi.org/10.1007/s42417-022-00566-0

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and Polosukhin, "Attention is all you need" *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 5998-6008, 2017. https://doi.org/10.48550/arXiv.1706.03762

[10] J. Xu, H. Wu, J. Wang, and M. Long, "Anomaly transformer: Time series anomaly detection with association discrepancy", in *Proc. Int. Conf. on Learning Representations (ICLR)*, 2022. https://doi.org/10.48550/arXiv.2110.02642

[11] N. Dutta, P. Kaliannan, U. Subramaniam, "Application of machine learning algorithm for anomaly detection for industrial pumps. Machine Learning Algorithms for Industrial Applications", *Springer*, pp. 237-263, 2021. https://doi.org/10.1007/978-3-030-50641-4_14

[12] Z. Yang, J. Cen, X. Liu, J. Xiong, and H. Chen, "Research on bearing fault diagnosis method based on transformer neural network", *Meas. Sci. Technol.*, vol. 33, no. 8, p. 085111, 2022. https://doi.org/10.1088/1361-6501/ac66c4

[13] S. Manikandan, and K. Duraivelu, "Fault diagnosis of various rotating equipment using machine learning approaches–A review", *Proc. Inst. Mech. Eng., Part E: J. Process Mech. Eng.*, vol. 235, no. 2, pp. 629-642, 2021. https://doi.org/10.1177/0954408920971976

[14] C. E. Sunal, V. Dyo, and V. Velisavljevic, "Review of machine learning based fault detection for centrifugal pump induction motors", *IEEE Access*, vol. 10, pp. 71344-71355, 2022. https://doi.org/10.1109/ACCESS.2022.3187718

[15] H. Pan, X. He, S. Tang, F. Meng, "An improved bearing fault diagnosis method using one-dimensional CNN and LSTM", *Stroj. Vestn./J. Mech. Eng.*, vol. 64, pp. 443-452, 2018. https://doi.org/10.5545/sv-jme.2018.5249.

[16] H. Sun, and S. Zhao, "Fault diagnosis for bearing based on 1DCNN and LSTM", *Shock Vib.*, pp.1-17, 2021. https://doi.org/10.1155/2021/1221462

[17] H. Wu, M. J. Triebe, and J. W. Sutherland, “A transformer-based approach for novel fault detection and fault classification/diagnosis in manufacturing: A rotary system application”, *J. Manuf. Syst.*, vol. 67, pp. 439-452, 2023. https://doi.org/10.1016/j.jmsy.2023.02.018

[18] Abdulkarem, Waleed, R. Amuthakkannan, and Khalid F. Al-Raheem, “Centrifugal pump impeller crack detection using vibration analysis”, in *Proc. Int. Conf. on Research in Science, Engineering and Technology*, 2014, pp. 206-211. http://dx.doi.org/10.15242/IIE.E0314606.

[19] W. Jiang , C. Wang , J. Zou, and S. Zhang, “Application of deep learning in fault diagnosis of rotating machinery”, *Processes*, vol. 9, no. 6, p. 919, 2021. https://doi.org/10.3390/pr9060919

[20] Q. Zhao, G. Cheng, X. Han, D. Liang, X. Wang, “Fault diagnosis of main pump in converter station based on deep neural network”, *Symmetry*, vol. 13, no. 7, p. 1284, 2021. https://doi.org/10.3390/sym13071284

[21] Y. Jin, L. Hou, and Y. Chen, “A time series transformer based method for the rotating machinery fault diagnosis”, *Neurocomputing*, vol. 494, pp. 379-395, 2022. https://doi.org/10.1016/j.neucom.2022.04.111

[22] Z. Lu, L. Liang, J. Zhu, W. Zou, and L. Mao, “Rotating machinery fault diagnosis under multiple working conditions via a time series transformer enhanced by convolutional neural network”, *IEEE T. Instrum. Meas.*, vol. 72, pp. 1-11, 2023. https://doi.org/10.1109/TIM.2023.3318707

[23] J. Vanga, D. P. Ranimekhala, S. Jonnala, J. Jamalapuram, B. Gutta, S. R. Gampa, and A. Alluri, “Fault classification of three phase induction motors using Bi-LSTM networks”, *J. Electr. Syst. Inf. Technol.*, vol. 10, no. 1, pp. 1-15, 2023. https://doi.org/10.1186/s43067-023-00098-x

[24] J. P. S. Gonçalves, F. Fruett, J. G. Dalfré Filho, and M. Giesbrecht, “Fault detection and classification in a centrifugal pump from vibration data using markov parameters”, *Mech. Syst. Signal Process.*, vol. 158, p. 107694, 2021. https://doi.org/10.1016/j.ymssp.2021.107694

[25] A. Chabert, M. C. Bakkay, P. Schweitzer, S. Weber S, and J. Andrea, “A transformer neural network for AC series arc-fault detection”, *Eng. Appl. Artif. Intell.*, vol.125, p. 106651, 2023. https://doi.org/10.1016/j.engappai.2023.106651