# Hybrid ARIMAX–LSTM Modeling for Enhanced Vibration Prediction in Rotating Machinery: Application to a Cement Mill Fan

**Noureddine Allassak**
IPSS Laboratory, Computer Science Department, Faculty of Sciences, Mohammed V University in Rabat, Rabat, Morocco
noureddine.allassak@um5r.ac.ma (corresponding author)

**Salima Trichni**
IPSS Laboratory, Computer Science Department, Faculty of Sciences, Mohammed V University in Rabat, Rabat, Morocco | Department of Transversal Modules, Faculty of Law, Economics and Social Sciences, Mohammed V University in Rabat, Rabat, Morocco
s.trichni@um5r.ac.ma

**Fouzia Omary**
IPSS Laboratory, Computer Science Department, Faculty of Sciences, Mohammed V University in Rabat, Rabat, Morocco
f.omary@um5r.ac.ma

*Received: 30 April 2025 | Revised: 11 June 2025 | Accepted: 15 June 2025*

*DOI: https://doi.org/10.48084/etasr.11845*

## ABSTRACT

**In intensive industrial applications, such as cement manufacturing, the reliable operation of key equipment is critical to ensure the equipment longevity and a continuous efficient production. This study focuses on forecasting the mechanical vibrations in an Induced Draft (ID) fan used in a cement mill, where abnormal vibration levels may indicate impending faults or performance degradation. Given the dynamic and nonlinear behavior of such systems, accurate prediction is both challenging and essential for condition-based maintenance. A hybrid forecasting framework is utilized in the current study, that integrates the statistical accuracy of the Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX) model with the nonlinear learning capabilities of Long Short-Term Memory (LSTM) networks. The ARIMAX component captures the linear structure and exogenous influences, while the LSTM models capture the residual nonlinearities providing a holistic approach to the vibration behavior. The proposed hybrid model is benchmarked against three standalone approaches: ARIMAX, LSTM, and the Machine Learning (ML)-based XGBoost algorithm. The experimental results demonstrate that the hybrid ARIMAX–LSTM model significantly outperforms individual models in terms of prediction accuracy, as measured by the RMSE, MAE, and $R^2$ statistical metrics. These findings highlight the potential of combining classical time series models with Deep Learning (DL) architectures for advanced prognostics in industrial rotating machinery.**

***Keywords-predictive maintenance; LSTM; ARIMAX; DL; vibration; fan***

## I. INTRODUCTION

Monitoring and predicting the mechanical vibrations in industrial rotating machinery is essential to ensure reliability, prevent critical failures, and enable Predictive Maintenance (PdM) strategies [1]. In cement manufacturing plants, ID fans used in cement mill systems operate under harsh conditions and are subject to varying mechanical and thermal stresses [2]. Unexpected fluctuations in vibration levels can lead to wear, imbalance, or the onset of structural defects. Time series forecasting includes the presence of complex nonlinearities and the influence of exogenous operational variables, such as speed, airflow, and winding temperatures. Traditional linear models, such as the ARIMAX, have proven effective for capturing temporal dependencies and incorporating external inputs [3]. Nevertheless, their ability to model nonlinear behavior remains limited.

LSTM networks have emerged as a particularly powerful solution for time series data capturing intricate temporal and nonlinear relationships [4]. LSTM networks are well-suited to learning long-term dependencies and adapting to dynamic signal patterns, which are characteristic of industrial vibration signals. This study presents a hybrid modeling approach of ARIMAX and LSTM networks. The hybrid model first applies ARIMAX to model linear trends and exogenous effects, then feeds the residual errors into an LSTM model to capture the remaining nonlinear components. To validate the performance of this approach, a comprehensive comparison is conducted against standalone models, including ARIMAX, LSTM, and XGBoost [5].

PdM has the ability to anticipate equipment failures and optimize maintenance schedules. ML and DL suggest remarkable potential in capturing the complex patterns in sensor data for fault detection and life estimation [6]. In [7], XGBoost outperformed several ML models in classifying machine failures from a real-world PdM dataset, demonstrating its effectiveness even on unbalanced data. Meanwhile, authors in [8] used XGBoost as a robust classifier to fuse empirical and adaptively extracted features from a neural network, achieving high fault classification accuracy and resilience to noise. These findings confirm XGBoost's capability to serve as a reliable tool for PdM and intelligent fault diagnosis in complex industrial settings.

Authors in [9] developed a LightGBM-based framework for diagnosing faults in induction motors, presenting strong performance even under unseen operating conditions. Their method combines advanced feature selection using a modified recursive elimination process, robust validation via leave-one-loading-out cross-validation, and Bayesian optimization. Tests on real datasets yielded up to 100% accuracy, confirming the model's effectiveness.

In addition to classification tasks, regression techniques have been applied to estimate the Remaining Useful Life (RUL) of critical components. One study employed a PCA-SVR pipeline, where wavelet-based denoising is followed by feature extraction and dimensionality reduction. The SVR model trained on the reduced features is able to forecast the residual lifespan of bearings with high precision [10].

In [11], condition monitoring in industrial production relies heavily on manual vibration analysis, which is time-consuming and expertise dependent. The research proposes a semi-automated diagnostic system that uses Fast Fourier Transform (FFT) spectra to extract energy features and apply ML for fault detection. The system, utilizing Support Vector Machines (SVM) and ensemble algorithms, provides accurate diagnostics with minimal manual intervention and early warnings, achieving over 90% accuracy in real-world conditions testing.

Authors in [12] propose a novel approach for predicting the RUL of bearings, integrating a health indicator derived from cumulative modified multiscale permutation entropy with an LSTM model. A virtual health degree, modeled through an exponential degradation function, is introduced as the prediction target. The experimental results confirm the method's effectiveness, achieving better accuracy and lower error rates than traditional techniques, thus enhancing the reliability of the prognostic health management systems.

Collectively, these contributions underscore the growing trend of integrating ML and DL models into PdM systems. However, challenges, such as generalization to new environments, model interpretability, and robustness against noise, still demand further investigation, particularly in industrial contexts where operational variability is significant. The proposed hybrid architecture is applied to a real-world dataset collected from a high-capacity fan operating in a cement mill, with the results showing superior accuracy and robustness in vibration prediction compared to the aforementioned methods.

## II. METHODOLOGY

This section outlines the methodological framework adopted to develop the proposed hybrid forecasting model for vibration prediction. The approach integrates ARIMAX, a linear statistical model to extract the primary trends and linear dependencies from the data. In addition it employs LSTM, a DL model to capture the complex nonlinearities embedded in the residuals of the statistical model. The complete methodology is summarized in Figure 1.

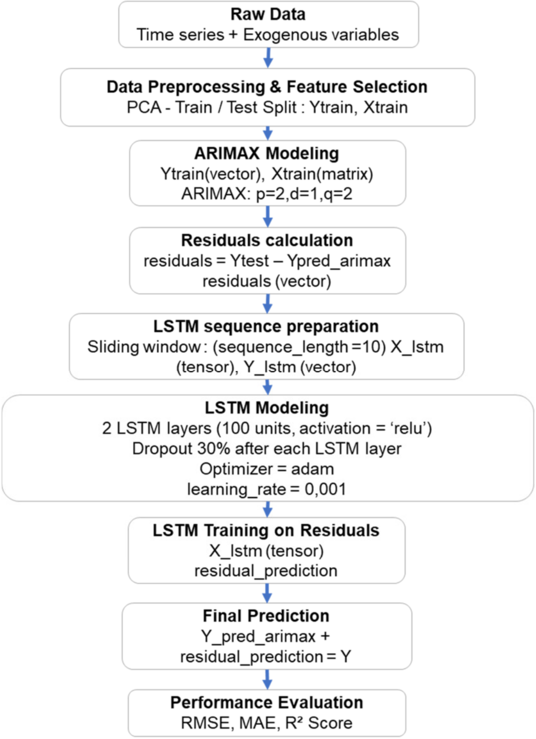


Fig. 1. Flowchart of the proposed ARIMAX-LSTM hybrid approach.

The experimental validation is conducted on a real-world dataset acquired from an ID fan operating in the grinding section of a cement plant, as depicted in Figure 2.

Positioned downstream of the vertical mill, the fan plays a pivotal role in maintaining the airflow and pressure balance, ensuring stable and energy-efficient milling processes [13, 14]. Due to its continuous operation under fluctuating load conditions, the ID fan is susceptible to mechanical degradation, making it an ideal candidate for PdM. The dataset consists of 18496 time-stamped entries recorded, encompassing sensor measurements, such as current, speed, component temperatures, airflow, and the target variable: vibration intensity.

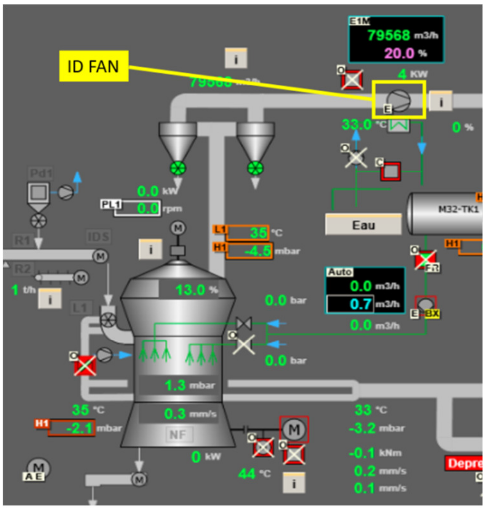


Fig. 2. Schematic representation of the ID fan used in the experimental setup.

Table I provides a comprehensive overview of the dataset's key variables, including their descriptions and measurement units as well as a statistical summary of the dataset.

### *A. Data Preprocessing and Feature Selection*

The preprocessing phase involved outlier and missing value removal in accordance with the knowledge and insights provided by plant experts, to ensure that the dataset was suitable for time series modeling.

Any entries with missing values in the target variable, vibration, are removed to maintain the prediction accuracy and model reliability. A total of 241 missing values were removed. To guarantee physical coherence, the outlier filtering retained only data points with vibration values between 0.4 mm/s and 7 mm/s and intensity values greater than 450 A to reflect operationally relevant scenarios. Finally, during the feature selection, to identify the most relevant variables for predicting the vibration levels Principal Component Analysis (PCA) is performed [15]. As illustrated in Figure 3, the first two principal components accounted for a substantial portion of the variance in the dataset, with the first component explaining 76.88% and the second 10.69%, resulting in a cumulative variance of 87.57%. This indicates that the majority of the information contained in the original feature space can be effectively represented in a lower-dimensional subspace.

TABLE I. DESCRIPTION AND STATISTICAL SUMMARY OF FEATURES

| Feature | Description | Unit | Mean | Standard Deviation | Min | Max |
|---|---|---|---|---|---|---|
| Intensity | Current drawn by the system | A | 853.13 | 449.94 | 0 | 1326.43 |
| Speed | Rotational speed of the equipment | RPM | 530.12 | 431.22 | 0 | 917.79 |
| Winding 1 Temperature | Temperature of electrical winding 1 | °C | 83.61 | 36.59 | 0 | 135.95 |
| Winding 2 Temperature | Temperature of electrical winding 2 | °C | 81.62 | 35.42 | 0 | 132.92 |
| Winding 3 Temperature | Temperature of electrical winding 3 | °C | 83.48 | 34.25 | 0 | 136.84 |
| Bearing 1 Motor Temperature | Temperature of motor bearing 1 | °C | 51.77 | 19.33 | 0 | 92.46 |
| Bearing 2 Motor Temperature | Temperature of motor bearing 2 | °C | 58.23 | 23.62 | 0 | 99.01 |
| Vibration | Vibration intensity measured | mm/s | 1.86 | 1.77 | 0 | 37.00 |
| Bearing 1 Fan Temperature | Temperature of fan bearing 1 | °C | 35.92 | 13.7 | 0 | 75.04 |
| Bearing 2 Fan Temperature | Temperature of fan bearing 2 | °C | 36.23 | 13.56 | 0 | 71.7 |
| Air Flow | Airflow rate within the system | m³/h | 122234.18 | 100359.31 | 0 | 496975.59 |

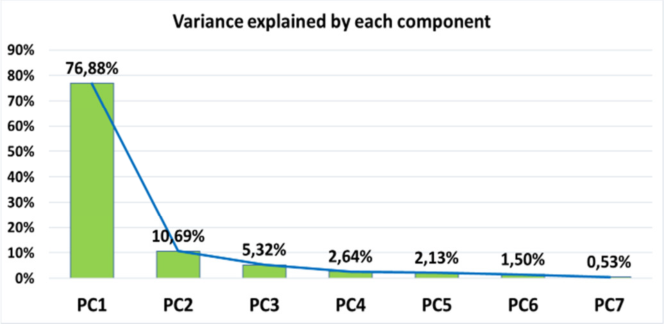


Fig. 3. Variance explained by each principal component during PCA.

Based on the feature loadings on these components, the following variables are identified as the most influential: speed, winding 1 temperature, winding 2 temperature, air flow, winding 3 temperature, bearing 2 motor temperature, and bearing 1 motor temperature, as illustrated in Figure 4.

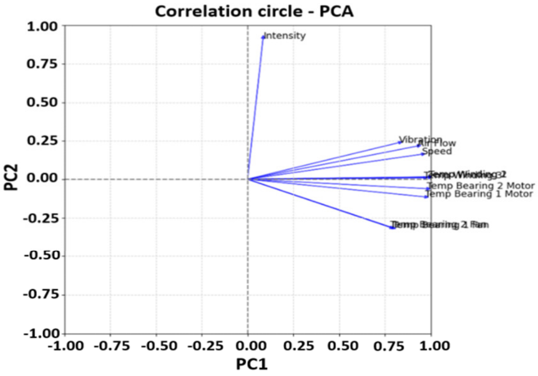


Fig. 4. PCA correlation circle.

### *B. ARIMAX Modeling*

An ARIMAX model is employed to model the linear temporal dynamics of the vibration signal while accounting for the influence of exogenous variables. The ARIMAX framework extends the traditional ARIMA model by incorporating additional covariates.

### *1) Stationarity Assessment*

The Augmented Dickey-Fuller (ADF) test was conducted to assess the stationarity of the vibration series [16]. The results are presented in Table II.

TABLE II. ADF TEST RESULTS

| ADF Statistic | p-value | Critical Values | | |
|---|---|---|---|---|
| | | 1% | 5% | 10% |
| -5.64 | 0.00001 | -3.43 | -2.86 | -2.57 |

The test yielded a statistic of -5.64, which is significantly lower than the 1%, 5%, and 10% critical values (-3.43, -2.86, -2.57, respectively), with a p-value smaller than 0.00001. This strongly suggests that the time series is stationary after first-order differencing, thereby justifying the choice of $d$=1.

### *2) Model Selection via AIC/BIC*

The present study evaluated several candidate ARIMAX models with $p,q \in [0,2]$ and fixed d=1, computing both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for model comparison. The results in Table III and Figure 5 suggest that the ARIMAX(2,1,2) model consistently achieved the lowest AIC (9743.98) and BIC (9781.2) scores, outperforming alternative configurations, such as ARIMAX(1,1,2), ARIMAX(2,1,1), and ARIMAX(1,1,1). This indicates a better trade-off between the model fit and complexity.

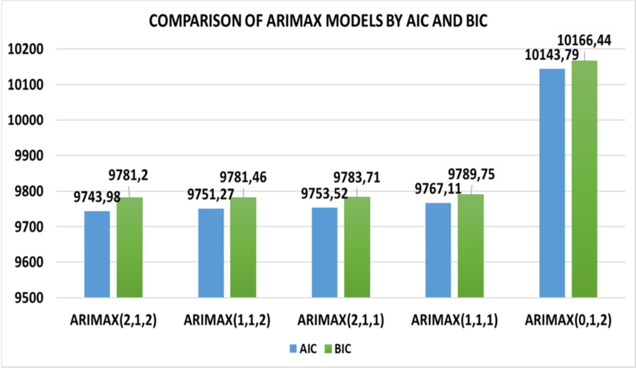


Fig. 5. ARIMAX models ranked by AIC.

TABLE III. COMPARISON OF ARIMAX MODELS BY AIC AND BIC

| Model | AIC | BIC |
|---|---|---|
| ARIMAX(2,1,2) | 9743.98 | 9781.2 |
| ARIMAX(1,1,2) | 9751.27 | 9781.46 |
| ARIMAX(2,1,1) | 9753.52 | 9783.71 |
| ARIMAX(1,1,1) | 9767.11 | 9789.75 |
| ARIMAX(0,1,2) | 10143.79 | 10166.44 |

### *3) Autocorrelation Diagnostic*

The Autocorrelation Function (ACF) of the differenced series exhibited slow decay beyond lag 1, while the Partial Autocorrelation Function (PACF) demonstrated significant spikes at lags 1 and 2. This pattern aligns with an autoregressive component of order 2, thus justifying p=2 and q=2. Subsequently, ARIMAX(2,1,2) is the most robust and parsimonious model for capturing the linear dependencies in the vibration signal. The residuals from this model, which encapsulate the unexplained non-linear components, were then used as an input to the LSTM network in the proposed hybrid forecasting architecture.

## *C. Residual Computation and Motivation for LSTM Modeling*

To capture the nonlinear components in the residuals, an LSTM network with two hidden layers is employed, each comprising 100 units. Dropout layers are inserted between the LSTM layers to mitigate overfitting, with a dropout rate of 0.3. The model is compiled using Mean Squared Error as the loss function and the Adam optimizer with a Learning Rate (LR) of 0.001. Training is conducted over 100 epochs with a batch size of 32, and early stopping is enabled. The hyperparameters of the LSTM model are identified through a comprehensive Grid Search, which systematically assessed all potential combinations of the parameter values, as presented in Table IV.

TABLE IV. GRID SEARCH HYPERPARAMETER COMBINATIONS FOR LSTM

| Hyperparameter | Values |
|---|---|
| Number of layers | 1, 2 |
| Batch size | 16, 32, 64, 128 |
| Optimizer | Adam |
| Units | 40, 50, 60, 100 |
| LR | 0.001, 0.0005 |
| Epochs | 20, 30,50, 100 |

This rigorous approach facilitated the optimal extraction of temporal features. The selected optimal hyperparameters resulting from this process are summarized in Table V.

TABLE V. BEST HYPERPARAMETERS FOR LSTM MODEL

| Hyperparameter | Value |
|---|---|
| Layers | 2 |
| Units per Layer | 100 |
| Dropout Rate | 0.3 |
| LR | 0.001 |
| Batch Size | 32 |
| Epochs | 100 |

The loss function curve depicted in Figure 6 offers valuable insights into the model's training dynamics [17]. As anticipated, the curve exhibits a consistent decline in loss across the training epochs, reflecting the LSTM model's effectiveness in minimizing the error and progressively enhancing its performance.

This steady reduction indicates that the selected hyperparameters support an optimally balanced training process, successfully avoiding both underfitting and overfitting. Moreover, the smooth trajectory of the curve underscores the model's stability and reliability throughout the training phase.

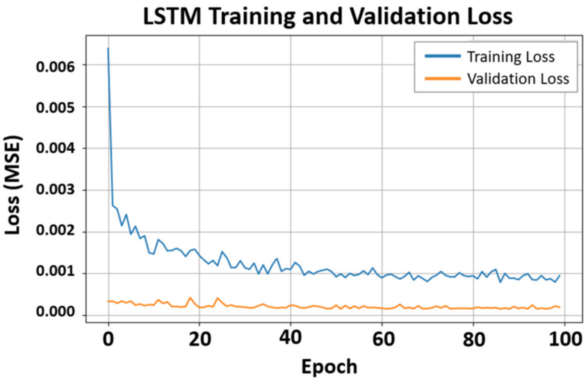


Fig. 6. LSTM loss function curve.

### D. *Construction of the Hybrid Forecast*

After training, the LSTM model is used to predict the residuals over the test set. These predicted residuals are then transformed back to their original scale via the inverse transformation of the Min-Max-Scaler [18, 19]. The final hybrid prediction is obtained by summing the ARIMAX predictions and the LSTM-predicted residuals, as shown in:

$$\hat{y}_t^{\,Hybrid} = \hat{y}_t^{\,ARIMAX} + \hat{y}_t^{\,LSTM} \tag{1}$$

### E. *Evaluation Metrics*

To assess the performance of the proposed hybrid ARIMAX-LSTM model, three standard evaluation metrics are utilized: The Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE), and the Coefficient of Determination ($R^2$) [20]. These metrics provide complementary insights into the accuracy and robustness of the predictions.

MAE measures the average magnitude of the errors between the predicted and the actual values, without considering their direction. It is defined in:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{2}$$

where $n$ is the number of observations, $y_i$ is the true value, and $\hat{y}$ is the predicted value. A lower MAE indicates better model accuracy.

RMSE measures the square root of the average of squared differences between the predicted and actual values. It penalizes larger errors more heavily, thus emphasizing significant deviations. Lower RMSE values denote higher predictive performance. It is given by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

The $R^2$ score quantifies the proportion of the variance in the dependent variable that is predictable from the independent variables [21]. It is computed by:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2} \tag{4}$$

where $\bar{y}$ represents the mean of the actual values. An $R^2$ score closer to 1 indicates a strong explanatory power of the model. Together, these three metrics allow for a comprehensive evaluation of both the average and extreme predictive errors, as well as the overall goodness-of-fit of the model.

## III. RESULTS AND DISCUSSION

The predictive performance of the proposed hybrid ARIMAX–LSTM model is evaluated against three baseline models: standalone LSTM, XGBoost, and standalone ARIMAX. Table VI and Figure 7 summarize the comparison results.

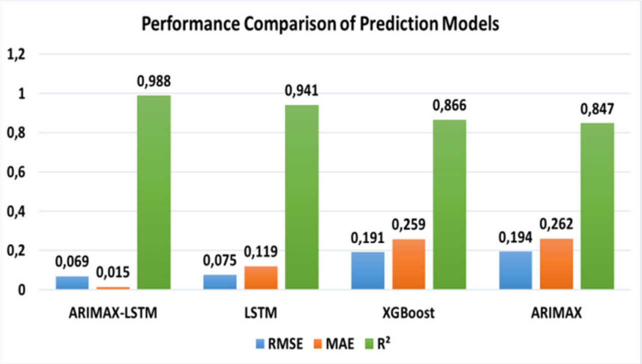


Fig. 7. Performance comparison of prediction models.

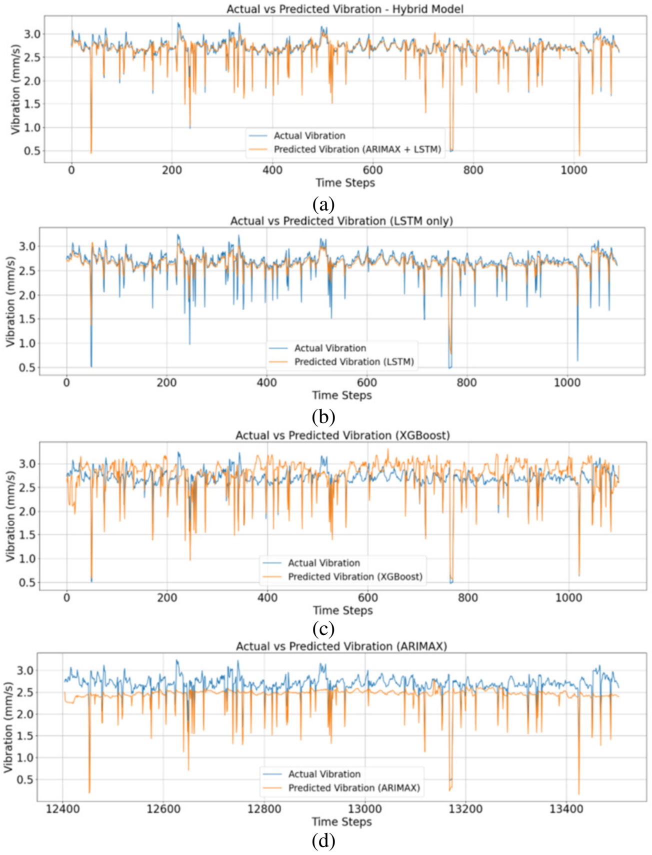
TABLE VI. PERFORMANCE COMPARISON OF PREDICTION MODELS BASED ON RMSE, MAE, AND $R^2$

| | **Metrics** | | |
|---|---|---|---|
| **Model** | **RMSE** | **MAE** | ***R²*** |
| ARIMAX-LSTM | 0.069 | 0.015 | 0.988 |
| LSTM | 0.075 | 0.119 | 0.941 |
| XGBoost | 0.191 | 0.259 | 0.866 |
| ARIMAX | 0.194 | 0.262 | 0.847 |

The hybrid ARIMAX–LSTM model consistently outperformed all other approaches across all evaluation metrics. It achieved an RMSE of 0.069 and an MAE of 0.015, both significantly lower than those of the other models. Moreover, its $R^2$ score of 0.988 indicates an exceptional ability to explain nearly all the variance in the vibration data, underscoring the model's robustness and reliability. Compared to the standalone LSTM, which already demonstrates a strong $R^2$ score of 0.941, the hybrid model provides a substantial improvement in both the error reduction and variance explanation. This confirms that explicitly modeling the linear components through ARIMAX before addressing the nonlinear residual patterns with LSTM yields a superior performance over purely data-driven approaches.

Furthermore, when comparing XGBoost and standalone ARIMAX with $R^2$ scores of 0.866 and 0.847, respectively, the advantages of the hybridization strategy are evident. XGBoost struggled to capture the temporal dependencies inherent in time series forecasting, resulting in larger errors. Similarly, ARIMAX failed to adequately model the complex nonlinearities present in the vibration signals.

Overall, the results substantiate that the hybrid ARIMAX–LSTM model effectively combines the strengths of both statistical and DL paradigms, leading to significant gains in prediction accuracy. Figures 8 illustrates the comparison between the actual and predicted vibration values for each of the four models, highlighting their respective predictive performance.



(a)

(b)

(c)

(d)

Fig. 8. Actual against (a) ARIMAX-LSTM, (b) LSTM, (c) XGBoost, and (d) ARIMAX predicted vibration.

## IV. CONCLUSION

This study focused on accurate vibration prediction in rotating machinery. Leveraging the strengths of both statistical and Deep Learning (DL) techniques, a hybrid Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX)– Long Short-Term Memory (LSTM) model is proposed. Firstly, it captures the linear patterns and exogenous influences through ARIMAX, and secondly, it models the nonlinear residual components using LSTM networks.

The methodology was applied to a real-world dataset acquired from an Induced Draft (ID) fan in a cement mill, representative of the complex operating conditions and noise-prone measurements. To assess the effectiveness of the proposed approach, extensive experiments were conducted comparing the hybrid model with standalone ARIMAX, LSTM, and XGBoost models. The results demonstrated the superior performance of the hybrid model across all evaluation metrics, such as MSE, MAE, and $R^2$, confirming its ability to model the intricate temporal dependencies more effectively than the conventional methods. Compared to previous studies that relied solely on either statistical or Machine Learning (ML) models, the current work provides a meaningful advancement by combining the interpretability of ARIMAX with the expressive power of LSTM. This fusion allows for a more holistic modeling of industrial time series, especially in scenarios involving both exogenous variables and nonlinearity.

In summary, the proposed ARIMAX–LSTM architecture offers a robust and accurate solution for industrial vibration forecasting, making it a promising candidate for Predictive Maintenance (PdM) applications. Future work will focus on extending this hybridization with attention mechanisms or encoder–decoder frameworks, as well as deploying the model in real-time monitoring systems to validate its operational reliability and adaptability in dynamic production environments.

## ACKNOWLEDGMENT

The authors express their sincere gratitude to all stakeholders who contributed to this project, with particular acknowledgment to the engineers and technicians of the cement plant where the research was conducted.

## REFERENCES

[1] M. Romanssini, P. C. C. de Aguirre, L. Compassi-Severo, and A. G. Girardi, "A Review on Vibration Monitoring Techniques for Predictive Maintenance of Rotating Machinery," *Eng*, vol. 4, no. 3, pp. 1797–1817, Jun. 2023, https://doi.org/10.3390/eng4030102.

[2] F. A. Rodrigues and I. Joekes, "Cement industry: sustainability, challenges and perspectives," *Environmental Chemistry Letters*, vol. 9, no. 2, pp. 151–166, Oct. 2010, https://doi.org/10.1007/s10311-010-0302-2.

[3] V. Rathi, *Introduction to Time Series Analysis*. Delhi, India: Educohack Press, 2025.

[4] Y. Yu, X. Si, C. Hu, and J. Zhang, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures," *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019, https://doi.org/10.1162/neco_a_01199.

[5] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, May 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.

[6] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. da P. Francisco, J. P. Basto, and S. G. S. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers & Industrial Engineering*, vol. 137, Nov. 2019, Art. no. 106024, https://doi.org/10.1016/j.cie.2019.106024.

[7] D. K. Yadav, A. Kaushik, and N. Yadav, "Predicting machine failures using machine learning and deep learning algorithms," *Sustainable Manufacturing and Service Economics*, vol. 3, Aug. 2024, Art. no. 100029, https://doi.org/10.1016/j.smse.2024.100029.

[8] J. Xie, Z. Li, Z. Zhou, and S. Liu, "A Novel Bearing Fault Classification Method Based on XGBoost: The Fusion of Deep Learning-Based Features and Empirical Features," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, Dec. 2020, Art. no. 3506709, https://doi.org/10.1109/TIM.2020.3042315.

[9] A. Nemat Saberi, A. Belahcen, J. Sobra, and T. Vaimann, "LightGBM-Based Fault Diagnosis of Rotating Machinery Under Changing Working Conditions Using Modified Recursive Feature Elimination," *IEEE Access*, vol. 10, pp. 81910–81925, Aug. 2022, https://doi.org/10.1109/ACCESS.2022.3195939.

[10] S. Li, M. Li, Z. Liu, and M. Li, "A Data-Driven Residual Life Prediction Method For Rolling Bearings," in *Proceedings of 12th Data Driven Control and Learning Systems Conference*, Xiangtan, China, Jul. 2023, pp. 1629–1633, https://doi.org/10.1109/DDCLS58216.2023.10166951.

[11] A. F. Khalil and S. Rostam, "Machine Learning-based Predictive Maintenance for Fault Detection in Rotating Machinery: A Case Study," *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13181–13189, Apr. 2024, https://doi.org/10.48084/etasr.6813.

[12] P. K. Sahu and R. N. Rai, "LSTM-based deep learning approach for remaining useful life prediction of rolling bearing using proposed C-MMPE feature," *Journal of Mechanical Science and Technology*, vol. 38, no. 5, pp. 2197–2209, May 2024, https://doi.org/10.1007/s12206-024-0402-8.

[13] P. Pareek and V. S. Sankhla, "Review on vertical roller mill in cement industry & its performance parameters," *Materials Today: Proceedings*, vol. 44, pp. 4621–4627, Jan. 2021, https://doi.org/10.1016/j.matpr.2020.10.916.

[14] Y. Wang *et al.*, "Air balancing method of multibranch ventilation systems under the condition of nonfully developed flow," *Building and Environment*, vol. 223, Aug. 2022, Art. no. 109468, https://doi.org/10.1007/s12273-024-1189-3.

[15] H. Abdi and L. J. Williams, "Principal component analysis," *WIREs Computational Statistics*, vol. 2, no. 4, pp. 433–459, Jun. 2010, https://doi.org/10.1002/wics.101.

[16] R. Mushtaq, "Augmented Dickey Fuller Test." Social Science Research Network, Preprint, Rochester, NY, Aug. 17, 2011, https://doi.org/10.2139/ssrn.1911068.

[17] Q. Wang, Y. Ma, K. Zhao, and Y. Tian, "A Comprehensive Survey of Loss Functions in Machine Learning," *Annals of Data Science*, vol. 9, no. 2, pp. 187–212, Apr. 2020, https://doi.org/10.1007/s40745-020-00253-5.

[18] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, Oct. 2011.

[19] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, "Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification," in *Proceedings of Third International Conference on Smart Systems and Inventive Technology*, Tirunelveli, India, Dec. 2020, pp. 729–735, https://doi.org/10.1109/ICSSIT48917.2020.9214160.

[20] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014, https://doi.org/10.5194/gmd-7-1247-2014.

[21] F. Ruggeri, R. S. Kenett, and F. W. Faltin, *Encyclopedia of Statistics in Quality and Reliability*. NJ, USA: Wiley, 2008.