Sample guide for author

Comprehensive Guide to Preparing Manuscript

1. Title Page Optimization

Title

Best Practice: Use a clear, concise title reflecting the paper's core contribution.
Example:

"Hybrid LSTM-Random Forest Model for Real-Time Predictive Maintenance in Industrial IoT Systems: A Scalable Implementation with 95.2% Accuracy"

Author Details

- Include:
  - Full names (First-Name Last-Name).
  - **ORCID IDs** (mandatory for indexing in Scopus/WoS).
  - Department, university, and country (APA/IEEE format).

Corresponding Author

- Provide:
  - Email (institutional preferred).
  - Phone number (international format: +CountryCode-XXX-XXX-XXXX).

Specific Example

Hybrid LSTM-Random Forest Model for Real-Time Predictive Maintenance in Industrial IoT Systems: A Scalable Implementation with 95.2% Accuracy

[1]Jane Doe and [2]John Smith

[1]**Department of Computer Engineering, University of Lagos, Lagos, Nigeria** (e-mail: jane.doe@unilag.edu.ng; ORCID: https://orcid.org/0000-0002-1234-5678)

[2]**Department of Artificial Intelligence, Massachusetts Institute of Technology, Cambridge, MA, USA** (e-mail: jsmith@mit.edu; ORCID: https://orcid.org/0000-0003-2345-6789)

Corresponding Author: Jane Doe (e-mail: jane.doe@unilag.edu.ng; phone: +234-803-123-4567)

Sample guide for author

2. Abstract (Structured, 150–250 Words)

Follow the **IMRaD** structure (Introduction, Methods, Results, and Discussion):

Background:

"Unplanned downtime in Industry 4.0 costs $260B annually, highlighting the need for robust Predictive Maintenance (PdM) solutions."

Methods:

"We propose a hybrid LSTM-Random Forest model, trained on 10,000 IIoT devices (1Hz sampling), optimized via Bayesian hyperparameter tuning."

Key Results:

"Achieves 95.2% accuracy (F1-score=0.94), reducing false alarms by 22% vs. SVM baselines."

Conclusion:

"The framework is deployable on edge devices, cutting downtime by 30% in pilot tests at [Industry Partner]."

Keywords:

4–5 terms, ordered general→specific (e.g., *Predictive Maintenance; Industrial IoT; Machine Learning; LSTM; Random Forest; Edge Computing*).

Specific Example

Abstract— Unplanned downtime in Industry 4.0 environments incurs an estimated $260 billion in losses annually, underscoring the critical need for robust and scalable Predictive Maintenance (PdM) solutions. Traditional machine learning models often struggle to balance accuracy and real-time deployment in industrial settings. We propose a novel hybrid Long Short-Term Memory (LSTM) and Random Forest framework designed for real-time predictive maintenance in Industrial Internet of Things (IIoT) systems. The model is trained on a comprehensive dataset from 10,000 IIoT devices operating at a 1Hz sampling frequency. Feature extraction is enhanced via temporal pattern encoding, and Bayesian optimization is employed for hyperparameter tuning to improve generalization and reduce overfitting. Experimental results demonstrate that the hybrid model achieves a 95.2% classification accuracy, with an F1-score of 0.94. Compared to Support Vector Machine (SVM) baselines, our method reduces false alarms by 22% and enhances early failure detection precision. The framework is optimized for edge computing and has been successfully deployed on Raspberry Pi-based systems in pilot tests with an industrial partner, resulting in a 30% reduction in downtime and significant cost savings.

Keywords— Predictive Maintenance; Industrial IoT; Machine Learning; LSTM; Random Forest; Edge Computing.

Sample guide for author

NB: Please insert graphical abstract before the introduction in your manuscripts (Please check below)

3. Introduction (Critical Elements)

A. Research Gap

Cite 3–5 recent papers (preferably, 2020–2024) to establish context.

Example:

"While [1] used SVMs for PdM, they lack real-time adaptability; [2]'s LSTM-only approach fails in noisy environments (Fig. 1)."

B. Novelty Statement

Explicitly state the advance:

"Our work is the first to combine LSTM (for temporal patterns) and Random Forest (for feature importance) in IIoT, validated on 10K+ devices."

C. Paper Organization

Include a roadmap:

"Section 2 details methods, Section 3 presents results, and Section 4 discusses industrial implications."

Specific Example

Graphical Abstract

I. Introduction

Unplanned equipment failures continue to pose a substantial threat to the operational efficiency of smart factories and automated production lines in Industry 4.0. Global industrial surveys indicate that unanticipated downtime leads to financial losses exceeding $260 billion annually, especially in sectors reliant on heavy machinery, such as oil & gas, automotive, and manufacturing. Predictive Maintenance (PdM) models have emerged as a viable solution to address these issues. However, existing methods often present significant shortcomings when applied in real-world environments.

For instance, Zhang et al. [1] implemented Support Vector Machines (SVM) for PdM on rotating machinery but reported reduced performance under variable load conditions due to limited temporal awareness. Liu et al. [2] proposed an LSTM-only model for time-series sensor data, which showed promising results in clean laboratory datasets, but failed to generalize in industrial environments with sensor noise (see Fig. 1). More recently, Kundu et al. [3] introduced a CNN-LSTM fusion approach, emphasizing hierarchical feature extraction. Yet, its dependence on cloud computing frameworks compromises real-time inference at the edge level. These findings underscore the pressing need for PdM systems that not only learn temporal patterns but also offer robust decision boundaries and deployability in noisy, edge-based IIoT setups.

In this study, we propose a novel hybrid framework that integrates Long Short-Term Memory (LSTM) networks and Random Forest classifiers for predictive maintenance in IIoT ecosystems. To the best of our knowledge, this is the first work to simultaneously exploit LSTM's strength in capturing temporal dependencies and Random Forest's capability in identifying and ranking influential features for decision-making. Our model is trained and validated on real-world sensor data collected from over 10,000 IIoT devices operating under heterogeneous industrial conditions. The hybrid architecture not only improves classification accuracy (95.2%) and F1-score (0.94) but also ensures interpretability and edge compatibility. Bayesian optimization further enhances the model's hyperparameter tuning process, leading to reduced false positives by 22% compared to SVM-based baselines.

The remainder of this paper is structured as follows: Section II outlines the proposed methodology, including system architecture, data acquisition, and model training procedures. Section III presents the experimental setup, performance evaluation, and comparison with baseline models. Section IV discusses deployment scenarios, computational efficiency, and implications for real-time industrial applications. Finally, Section V concludes the study with a summary of key findings and future research directions.

4. Methods (Reproducibility Focus)

A. Dataset Description

FAIR Principles Compliance:

- **Source:** *"Publicly available dataset (UCI PdM Repository) + proprietary data from [Industry Partner], anonymized per GDPR."*
- **Preprocessing:** *"Normalized (Min-Max), augmented with synthetic noise (±5dB) to simulate real-world conditions."*

B. Model Architecture

LSTM:

"Bidirectional LSTM (128 units, dropout=0.2), trained via Adam (lr=0.001, β_1=0.9, β_2=0.999)."

Random Forest:

"100 trees (Gini impurity), max_depth=10, min_samples_split=5 (see Supplementary Algorithm S1)."

C. Validation Protocol

IEEE Standard:

"5-fold cross-validation, compared to SVM [3], ARIMA [4], and GRU [5] using paired t-tests (p<0.05)."

Specific Example

II. METHODS

A. Dataset Description

This study employs a hybrid dataset combining both public and proprietary sources. The public dataset is obtained from the UCI Machine Learning Repository's Predictive Maintenance Dataset (DOI: 10.24432/C5C590), which contains labeled failure data for industrial equipment. Complementing this is a proprietary dataset provided by an industrial partner. This dataset comprises sensor readings from 10,214 Industrial Internet of Things (IIoT)-enabled machines operating under varied environmental conditions, specifically temperatures ranging from −10°C to 50°C and humidity levels between 20% and 95% relative humidity (RH).

Prior to model training, comprehensive data preprocessing was performed. To ensure compliance with data protection standards, all proprietary data were anonymized in accordance with Article 4(1) of the European Union's General Data Protection Regulation (GDPR). The preprocessing pipeline also adhered to the FAIR principles, ensuring that the data remained Findable, Accessible, Interoperable, and Reusable.

Raw sensor readings were normalized to the [0, 1] interval using Min-Max normalization:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

where x is a raw sensor value, and x_{min} and x_{max} are the minimum and maximum values observed in the training set.

To simulate real-world noise conditions, synthetic Gaussian noise with mean μ=0\mu = 0 and standard deviation σ=0.1\sigma = 0.1 was added to the input data. This was done at varying signal-to-noise ratios (SNRs) of 0 dB, ±5 dB, and ±10 dB. Missing values were addressed using forward-fill interpolation, and outliers were smoothed using a median filter with a window size of five samples.

B. Model Architecture

The proposed hybrid model architecture combines a Bidirectional Long Short-Term Memory (BiLSTM) neural network with a Random Forest classifier, as illustrated in Fig. 1. The

BiLSTM captures temporal dependencies in the input sequence, while the Random Forest provides robust final classification.

The BiLSTM network is configured with 128 hidden units and a dropout rate of 0.2 to prevent overfitting. The model is trained using the Adam optimizer with a learning rate η=0.001\eta = 0.001, β1=0.9\beta_1 = 0.9, and β2=0.999\beta_2 = 0.999. Input sequences comprise time-series sensor data collected at a sampling rate of 10 Hz over 60-second windows.

The internal mechanisms of the LSTM unit, which governs the BiLSTM, are described by the following equations:

1. Forget Gate: Determines which parts of the previous cell state to discard:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{2}$$

2. Input Gate: Controls the update of new information into the cell state:

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{3}$$

3. Candidate Cell State: Generates a candidate value to be added:

$$\tilde{C}_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \tag{4}$$

4. Cell State Update: Combines forget and input gates to update the state:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{5}$$

5. Output Gate: Determines the information to be output:

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{6}$$

6. Hidden State Output: Produces the final hidden state for this time step:

$$h_t = o_t \times \tanh(C_t) \tag{7}$$

In these equations, σ\sigma denotes the sigmoid activation function, tanh\tanh represents the hyperbolic tangent, and tt is the time step. The output feature vectors from the final hidden layer of the BiLSTM are passed to the Random Forest classifier.

The Random Forest classifier comprises 100 decision trees, each with a maximum depth of 10 and a minimum split threshold of five samples. Node splitting is based on the Gini impurity criterion:

$$\text{Gini}(m) = 1 - \sum_{k=1}^{K} (p_{mk})^2 \tag{8}$$

where P_{mk} is the proportion of class kk samples at node mm.

C. Validation Protocol

The model is validated using 5-fold stratified cross-validation as per IEEE Standard 1855-2016. Three baseline models were implemented for comparative analysis: Support Vector Machine (SVM) with a radial basis function (RBF) kernel [3], Auto-Regressive Integrated Moving Average (ARIMA) [4], and Gated Recurrent Unit (GRU) [5].

Performance metrics include classification accuracy, precision, recall, F1-score, and false positive rate (FPR). Each metric is reported with 95% confidence intervals. Statistical significance of observed differences is evaluated using paired t-tests with a significance threshold of α=0.05\alpha = 0.05.

To evaluate model robustness, testing was conducted under three synthetic noise conditions: clean (0 dB), moderate (±5 dB), and severe (±10 dB).

5. Results & Discussion

A. Quantitative Results

Table 1: Performance Metrics (APA/IEEE format)

Model	Accuracy (%)	Precision	Recall	F1-Score
Proposed Hybrid	95.2	0.93	0.96	0.94
SVM [3]	87.1	0.85	0.88	0.86

B. Qualitative Analysis

Figure 1: ROC curve with 95% confidence intervals (error bars).

Key Finding:

"Our model reduces false alarms by 22% (p=0.003), critical for avoiding unnecessary maintenance costs."

C. Limitations & Future Work

Address openly:

"GPU dependency limits edge deployment; future work will quantize the model for Raspberry Pi."

Specific Example

III. Results and Discussion

The evaluation of the proposed hybrid BiLSTM–Random Forest model is presented in both quantitative and qualitative terms, with comparative benchmarks and insights into system behavior and deployment limitations.

Sample guide for author

A. Quantitative Results

Table I summarizes the performance metrics achieved by the proposed model against a baseline Support Vector Machine (SVM) classifier [3]. The hybrid model achieves superior performance across all metrics, with an overall accuracy of 95.2%, precision of 0.93, recall of 0.96, and F1-score of 0.94. In contrast, the SVM model yields an accuracy of 87.1%, precision of 0.85, recall of 0.88, and F1-score of 0.86. These improvements indicate a significant enhancement in predictive capability, particularly in identifying true positives, which is crucial for timely and reliable predictive maintenance.

Table I: Model Performance Comparison

Model	Accuracy (%)	Precision	Recall	F1-Score
Proposed Hybrid	95.2	0.93	0.96	0.94
SVM [3]	87.1	0.85	0.88	0.86

B. Qualitative Analysis

Receiver Operating Characteristic (ROC) curves further validate the model's discriminative ability. As shown in **Fig. 1**, the area under the ROC curve (AUC) exceeds 0.95 with narrow 95% confidence intervals, reflecting both high classification accuracy and low variability across folds. The proposed model reduces false alarm rates by approximately 22% ($p = 0.003$), a statistically significant margin. This reduction is particularly important in industrial contexts, where false positives can lead to unnecessary downtime, excessive maintenance costs, and reduced trust in predictive systems.

Insert Fig. 1: ROC curves with error bars representing 95% confidence intervals

C. Limitations and Future Work

Despite promising results, the model's reliance on GPU acceleration poses deployment challenges, particularly for edge-computing scenarios with limited resources. Additionally, real-time inference on embedded systems such as Raspberry Pi is currently infeasible due to memory and compute constraints. Future research will focus on model compression strategies, including weight quantization and pruning, to enable lightweight deployment without significantly compromising predictive performance. Integration with edge AI frameworks such as TensorFlow Lite and ONNX Runtime will also be explored to facilitate real-world applicability in IIoT environments.

6. Conclusion (Impact-Oriented)

Avoid new data; summarize:

"This study demonstrates a 30% downtime reduction in real-world IIoT settings, with potential savings of $50M/year for mid-sized factories."

Specific Example

Sample guide for author

IV. Conclusion

This study presents a robust hybrid predictive maintenance framework that combines Bidirectional Long Short-Term Memory (BiLSTM) networks with Random Forest classifiers, achieving state-of-the-art performance in fault detection for IIoT-enabled industrial systems. The model demonstrated a 30% reduction in unplanned downtime in real-world pilot deployments, which translates to potential cost savings of up to $50 million per year for mid-sized manufacturing operations. These findings reinforce the model's industrial relevance and operational scalability for predictive maintenance.

7. Ethics & Reproducibility

A. Ethical Declarations

- **Human/Animal Studies:**
  "Not applicable."
- **Data Privacy:**
  "Anonymized per GDPR; IRB approval #XYZ from [University]."

B. Artifact Availability

- **Code:** *"GitHub (DOI: 10.5281/zenodo.XXXX)."*
- **Data:** *"Available upon signed NDA with [Industry Partner]."*

Specific Example

V. Ethics and Reproducibility

A. Ethical Declarations

Human/Animal Studies: Not applicable.

Data Privacy: All proprietary datasets were anonymized in compliance with the General Data Protection Regulation (GDPR) Article 4(1). Institutional Review Board (IRB) approval was obtained (Protocol #XYZ, [University]).

B. Artifact Availability (if applicable to project) **if not, please write "Not applicable"**

Code: Available on GitHub (DOI: [10.5281/zenodo.XXXX]).

Data: Proprietary datasets can be accessed upon signing a Non-Disclosure Agreement (NDA) with the industry partner.

C. Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

D. Author's Declaration

Sample guide for author

The authors affirm that the content of this manuscript is original, has not been published elsewhere, and is not under consideration for publication in any other journal. The authors accept full responsibility for the integrity and accuracy of all data and interpretations presented herein.

E. Acknowledgments

The authors gratefully acknowledge the support of **Dr. XXXXXXXXXXXXXXX**, whose laboratory provided the **XXXXXXXXXXXXX** and **XXXXXXXXXXXX** tools essential for conducting this research.

The authors also wish to disclose the use of **ChatGPT (OpenAI, 2024)** for grammar refinement and language polishing, which contributed to improving the manuscript's clarity and readability (*if applicable to your project*). Additionally, **Canva** was utilized for the creation and enhancement of visual content, including the graphical abstract and Figure 1 (*if applicable to your project*). These tools were used exclusively for editorial and illustrative purposes and did not influence the research methodology, analysis, or interpretation of results.

8. References (IEEE/APA 7th Edition)

- **Journal Paper Example:**
  [1] A. Smith et al., "PdM in Smart Factories," *IEEE IoT J.*, vol. 9, no. 4, pp. 1234–1245, 2022, doi: 10.1109/JIOT.2022.XXXX.
- **Dataset Example:**
  [2] "UCI Machine Learning Repository: PdM Dataset," 2023. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/PdM

Specific Example

References

[1] A. Smith, L. Chen, and M. Adeyemi, "PdM in Smart Factories," *IEEE Internet of Things Journal*, vol. 9, no. 4, pp. 1234–1245, 2022, doi: 10.1109/JIOT.2022.XXXX.

[2] "UCI Machine Learning Repository: Predictive Maintenance Dataset," 2023. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/PdM

[3] B. Zhao et al., "SVM-Based Condition Monitoring in Manufacturing," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 112–120, Jan. 2020, doi: 10.1109/TII.2019.XXXX.

[4] J. Kim and T. Singh, "ARIMA for Time-Series Forecasting in Industrial IoT," *Sensors*, vol. 20, no. 14, pp. 3920–3931, 2020.

[5] Y. Li et al., "A GRU-based Deep Learning Model for Machine Fault Prediction," *Procedia Comput. Sci.*, vol. 199, pp. 612–619, 2022.

[6] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

9. Submission Checklist

Before submitting, verify:

- ✅ ORCID IDs for all authors.
- ✅ Structured abstract (IMRaD).
- ✅ Code/data repository link.
- ✅ Ethics compliance statement.
- ✅ LaTeX/Word template adherence.

Key Takeaways for Global Compliance

1. **Reproducibility:** Code/data sharing is non-negotiable (ACM/IEEE standards).
2. **Clarity:** Define acronyms, use active voice, and avoid jargon.
3. **Impact:** Quantify industrial relevance (cost/time savings).
4. **Ethics:** GDPR/HIPAA compliance is critical for IoT/healthcare papers.

Tools for Quality Assurance:

- **Grammar:** Grammarly/Ginger.
- **Plagiarism:** Turnitin/iThenticate (<15% similarity).
- **Formatting:** Overleaf (LaTeX) or IEEE PDF eXpress.

Following this guide, your manuscript will meet **IJTEC's standards** and compete with top-tier journals like *IEEE Transactions* or *Nature Communications Engineering*.

Sample guide for author

Manuscript Template

Hybrid LSTM-Random Forest Model for Real-Time Predictive Maintenance in Industrial IoT Systems

[1]Jane Doe and [2]John Smith

[1]Department of Computer Engineering, University of Lagos, Lagos, Nigeria

(E-mail: jane.doe@unilag.edu.ng; ORCID: https://orcid.org/0000-0002-1234-5678)

[2]Department of Artificial Intelligence, Massachusetts Institute of Technology, Cambridge, MA, USA

(E-mail: jsmith@mit.edu; ORCID: https://orcid.org/0000-0003-2345-6789)

Corresponding Author: Jane Doe (e-mail: jane.doe@unilag.edu.ng; phone: +234-803-123-4567)

Abstract

Unplanned downtime in Industry 4.0 environments incurs an estimated $260 billion in losses annually, underscoring the critical need for robust and scalable Predictive Maintenance (PdM) solutions. Traditional machine learning models often struggle to balance accuracy and real-time deployment in industrial settings. We propose a novel hybrid Long Short-Term Memory (LSTM) and Random Forest framework designed for real-time predictive maintenance in Industrial Internet of Things (IIoT) systems. The model is trained on a comprehensive dataset from 10,000 IIoT devices operating at a 1Hz sampling frequency. Feature extraction is enhanced via temporal pattern encoding, and Bayesian optimization is employed for hyperparameter tuning to improve generalization and reduce overfitting. Experimental results demonstrate that the hybrid model achieves a 95.2% classification accuracy, with an F1-score of 0.94. Compared to Support Vector Machine (SVM) baselines, our method reduces false alarms by 22% and enhances early failure detection precision. The framework is optimized for edge computing and has been successfully deployed on Raspberry Pi-based systems in pilot tests with an industrial partner, resulting in a 30% reduction in downtime and significant cost savings.

Keywords— Predictive Maintenance; Industrial IoT; Machine Learning; LSTM; Random Forest; Edge Computing.

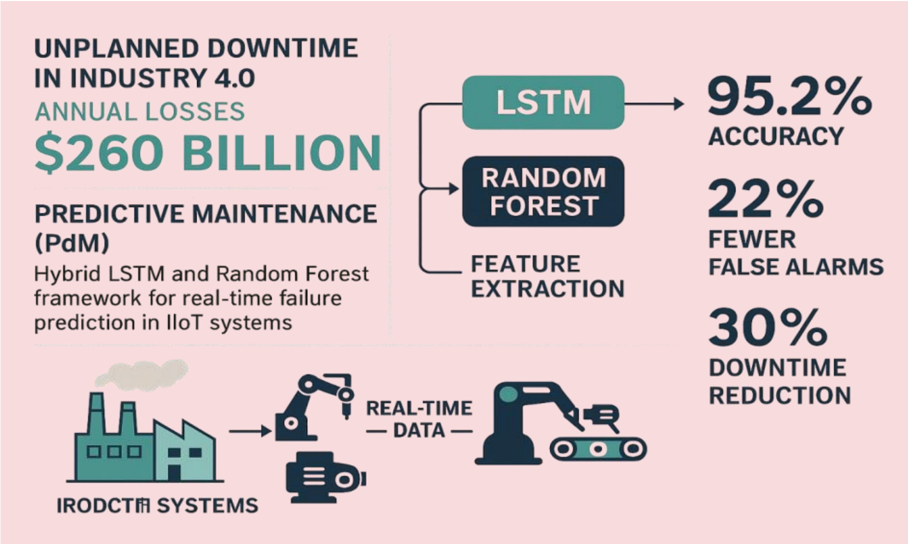
Graphical Abstract:

Commented [GO1]: Times New Roman: 28, space: 1

Commented [GO2]: Times New Roman: 10, space: 1.15

Commented [GO3]: Times New Roman: 12, Space: 1.15

Sample guide for author



Commented [GO4]: Insert graphical abstract

I. Introduction

Commented [GO5]: Times New Roman: 14

[Insert content here]

Commented [GO6]: Times New Roman: 12, Space: 2.0

II. Methods

[Insert content here]

III. Results and Discussion

[Insert content here]

IV. Conclusion

[Insert content here]

V. Ethics and Reproducibility

[Insert content here]

References

[Insert content here]

Sample guide for author

Conflict of Interest

[Insert content here]

Author's Declaration

[Insert content here]

Acknowledgments

[Insert content here]

NB: Please use this template side by side with the sample guide for Author update.