Enhancing machine failure prediction with a hybrid model approach

Ouiam Khattach[1,2], Omar Moussaoui[1], Mohammed Hassine[2,3]
[1]Mathematics, Signal and Image Processing, and Computing Research Laboratory (MATSI), Higher School of Technology (ESTO), Mohammed First University, Oujda, Morocco
[2]YosoBox, Oujda, Morocco
[3]Tisalabs Limited, Cork, Ireland

Article Info

Article history:

Received Jan 31, 2024
Revised Feb 20, 2024
Accepted Feb 28, 2024

Keywords:

Failure prediction
Hybrid LSTM–SVM
Internet of things
Long short-term memory
Support vector machine

ABSTRACT

The industrial sector is undergoing a substantial transformation by embracing predictive maintenance approaches, aiming to minimize downtime and reduce operational expenses. This transformative shift involves the incorporation of machine learning techniques to refine the accuracy of predicting machinery failures. In this article, we delve into an in-depth exploration of machine failure prediction, employing a hybrid model amalgamating long short-term memory (LSTM) and support vector machine (SVM). Our comprehensive study meticulously assesses the hybrid model's performance, comparing it with standalone LSTM and SVM models across three distinct datasets. The results showcase that the hybrid model outperformed, providing the modest dependable, and highest F1-score values in our evaluation.

Corresponding Author:

Ouiam Khattach
Mathematics, Signal and Image Processing, and Computing Research Laboratory (MATSI)
Higher School of Technology (ESTO), Mohammed First University
Oujda, Morocco
Email: ouiam.khattach@ump.ac.ma

1. INTRODUCTION

In the rapidly evolving industrial landscape, optimizing maintenance strategies is crucial to minimizing disruptions and ensuring seamless operations in the internet of things (IoT) applications. The potential for diverse applications extends into virtually all aspects of daily life for individuals, institutions, and broader societal contexts [1]. The application of IoT spans a wide range of areas including smart cities, healthcare, smart agriculture and water management, retail and logistics, smart living, and smart environment [2], [3]. Predictive maintenance is a transformative maintenance strategy that utilizes data analysis, sensor inputs, and machine learning to proactively anticipate equipment failures, deviating from traditional reactive or scheduled methods [4], [5]. By scrutinizing patterns, trends, and anomalies in data, predictive maintenance strives to forecast issues before they arise, representing a departure from conventional practices [6], [7]. The convergence of machine learning and predictive maintenance has paved the way for a paradigm shift in how organizations tackle equipment failures. Industries are progressively adopting predictive maintenance strategies, leveraging advanced technologies like machine learning to optimize maintenance schedules, enhance equipment reliability, and reduce operational expenses [8]. This shift underscores the transformative potential of data-driven strategies in redefining industrial maintenance practices. In this context, machine learning techniques such as K-means, random forests (RF), artificial neural networks (ANN), and support vector machines (SVM), play pivotal roles in performing predictive maintenance [9], [7]. Combining long short-term memory (LSTM) and SVM in a hybrid model enhances predictive accuracy, capitalizing on LSTM's sequential analysis and SVM's classification prowess. This fusion offers a comprehensive approach to anticipate and mitigate machine failures in industrial settings, with profound implications for transforming industrial maintenance practices.

This paper provides a comprehensive analysis of three distinct models, showcasing notable improvements in prediction accuracy compared to conventional single models. The paper is organized as follows: section 2 introduces the models used in our study. Section 3 summarizes related works on failure prediction using machine learning and deep learning techniques with different datasets. Section 4 presents comprehensive experimental results, including an explanation of the public datasets, intricate insights into the proposed LSTM-SVM architecture, and evaluation of performance metrics. Section 5, we delve into a detailed examination of the results and conduct a comparative analysis. Finally, the conclusion and future perspectives of the paper with a comprehensive summary are presented in section 6.

2. BACKGROUND

2.1. Long short-term memory neural network

The LSTM model belongs to the category of recurrent neural networks (RNNs) and is designed to address the issue of long-term dependencies in sequential data. Unlike standard RNNs, which can only capture information from nearby time steps, LSTM can effectively retain and utilize historical information over extended periods. The LSTM model functions through several key components: the block input, input gate, forget gate, output gate, and block output. LSTM proves to be particularly well-suited for tasks involving time series predictions and a wide array of other problems that require the retention of temporal memory. These applications span diverse fields such as natural language processing (including sentiment analysis), image and video captioning, and computer vision (including text recognition). Furthermore, combining LSTM with other models in hybrid architectures can often yield optimal performance in various tasks [10], [11]. Figure 1 represents the architecture of the LSTM model.

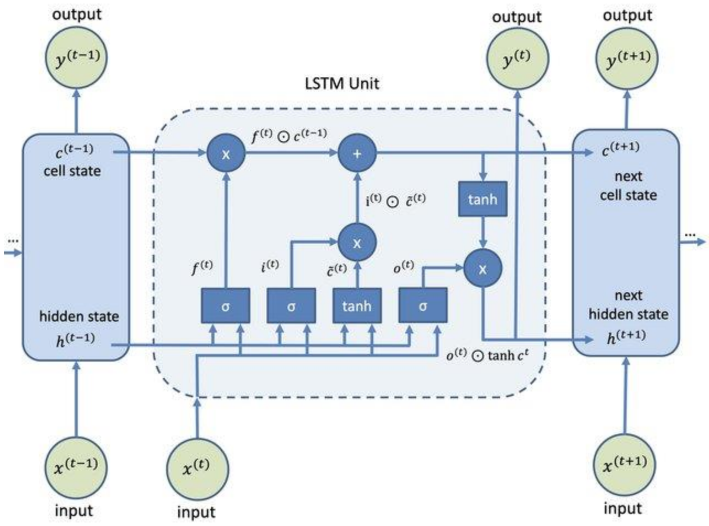


Figure 1. The architecture of the LSTM model

2.2. Support vector machine

The SVM is a prominent supervised machine learning algorithm firmly grounded in the principles of statistical learning theory, honed and refined over several decades. Its preeminence as a classification technique consistently positions it as a top performer, surpassing alternative methods with demonstrable superiority [12]. SVMs distinguish themselves in their capacity to oversee complex, high-dimensional datasets adeptly. This prowess stems from their intrinsic ability to discern the optimal hyperplane, which decisively demarcates data points among different classes while maximizing the margin, or separation distance, between them [13]. The versatility of SVMs extends to the realm of classification, where they exhibit competence in addressing both linear and non-linear challenges through the strategic application of diverse kernel functions. Consequently, SVMs find their application across an expansive spectrum of machine learning endeavors, encompassing tasks such as pattern recognition, object classification, and the nuanced domain of time series prediction and regression analysis. The unwavering robustness and adaptability inherent to SVMs solidify their status as an indispensable tool for practitioners in machine learning, ensuring their enduring relevance in navigating and mitigating the complexities inherent in real-world applications. Figure 2 illustrates the schematic architecture of the SVM classifier.

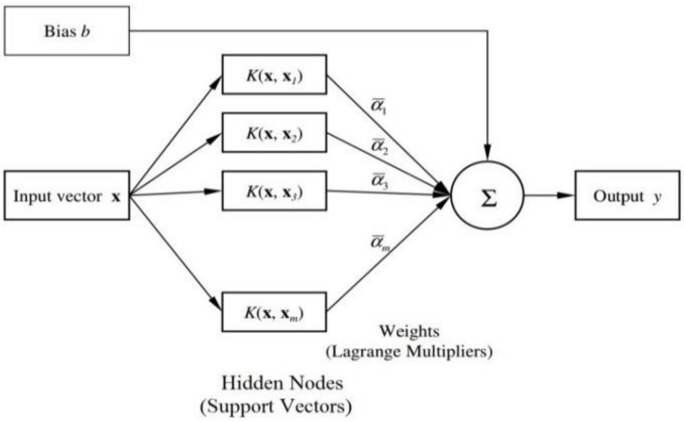


Figure 2. Schematic diagram of SVM architecture

3. RELATED WORK

The classification of failures in smart manufacturing has witnessed a surge in diverse methodologies, as extensively documented in the literature [8], [14]. The use of machine learning algorithms to identify irregular patterns in data, which are then used to enhance the accuracy and reliability of failure prediction models in various fields such as predictive maintenance and fault detection [15]. This survey provides an overview of machine learning and deep learning-based strategies for failure prediction.

Time series forecasting has seen the successful application of a variety of models, yet the task of choosing the suitable one remains a challenge. Recent years have seen a rise in interest in hybrid models, that combine various machine learning and deep learning techniques to address the complexities of failure prediction and time series forecasting [16]. These hybrid models leverage the strengths of different algorithms, enhancing their predictive power and robustness. This section explores the evaluation of hybrid models, their applications in real-world scenarios, and the potential benefits they offer in improving the reliability and efficiency of predictive maintenance and forecasting processes.

Wahid *et al.* [17] proposed a hybrid model for predictive maintenance based on a combination of convolutional neural networks (CNN) and LSTM. This model referred to as CNN-LSTM, utilized CNN for feature extraction from time series data and LSTM for prediction. The combined model outperformed individual ones, with CNN-LSTM exhibiting the highest prediction accuracy. This approach proved to be more reliable and suitable for predictive maintenance forecasting.

Borré *et al.* [18] proposed a CNN-LSTM hybrid model. The use of LSTM enabled the modeling of time series patterns, while CNN efficiently extracted vital features like trend changes and other commonly observed patterns in variable time series data. These findings offer substantial benefits to companies, enabling them to optimize maintenance schedules and enhance the overall performance of their electric machines.

Yeh *et al.* [19] suggested a hybrid network tailored to predict the extended maintenance duration of wind turbines, aiding efficient management in power companies. This model incorporated a combination of CNN and SVM. CNNs, adept at learning invariant features, were complemented by SVM, which excelled in producing precise decision surfaces when applied to well-structured feature vectors. The features extracted were then utilized as input to train a radial basis function support vector machine (RBF-SVM). The integration of these models resulted in remarkably high accuracy levels, showcasing the potential of combined approaches in predictive maintenance applications.

Vos *et al.* [20] address the challenge of fault detection in machine condition monitoring when there's limited faulty data for training. By using vibration signals from healthy systems and a combination of LSTM and SVM, the research successfully identifies abnormal mechanical behavior. The method works well for various scenarios, including gearbox tests and helicopter data, showcasing its effectiveness in machine condition monitoring, even when comprehensive faulty data is scarce.

In summary, this study highlights the growing importance of hybrid models in predicting failures in smart manufacturing. By combining machine learning and deep learning techniques, these hybrid models enhance prediction accuracy and reliability across domains like predictive maintenance and fault detection. These hybrid approaches hold significant promise in proving the efficiency and reliability of predictive maintenance.

4. PROPOSED METHOD

The objective of this work is to determine the most effective model to predict machine failure. We applied SVM, LSTM, and the hybrid model LSTM-SVM using three public datasets. To ensure a thorough assessment, we carefully isolate a subset from each dataset to construct an independent test set, separate from the training data. Additionally, a validation set is introduced to enhance the evaluation of the model's performance. This meticulous dataset division strategy aims to offer a robust assessment of the LSTM-SVM model's predictive capabilities.

4.1. Dataset description

4.1.1. Dataset

The dataset for this study pertains to predictive maintenance in industrial applications detailed in Table 1. The three datasets outlined in the table include context information, features extracted, and samples. These datasets are crucial for exploring and developing predictive maintenance models. They offer rich opportunities for analyzing anomalies and forecasting failures, key aspects of predictive maintenance.

Table 1. Datasets description

Data id	Datasets	Features extracted	Samples
Data 1	Distributed transformer monitoring	Oil temperature indicator, winding temperature indicator, ambient temperature indicator, oil level indicator, oil temperature indicator alarm, oil temperature indicator trip, magnetic oil gauge indicator	20465
Data 2	AnoML	Light, humidity, loudness, temperature	6558
Data 3	Predicting machine failure	Temperature, humidity, hours since previous failure, date.day-of-month, date.day-of-week, date-month, date-hour	8784

Firstly, data 1, known as the distributes transformer monitoring dataset, was meticulously collected through IoT devices over the period spanning from June 25th, 2019 to April 14th, 2020. This dataset stands out with a high-frequency update rate, providing new data points every 15 minutes and a total of 20,465 samples. It focuses on transformers, which hold a pivotal role in power systems, known for their reliability. However, they are susceptible to failure attributed to a multitude of internal and external factors. Among the various initiators of transformer failures, particularly mechanical and dielectric failures. Secondly, data 2 named the IoT anomaly detaction dataset, comprises 6,558 samples. This dataset is specifically curated for anomaly detection in IoT systems. This dataset encompasses a diverse array of components critical to IoT systems such as grove sensors, microcontrollers, shields, single-board computers, and an assortment of software tools. Lastly, data 3 serves as a foundational dataset for predicting failures in server machines operating within data center environments. This server machine failure prediction dataset comprises 8,784 samples, each contributing valuable data points to enhance predictive insights regarding server machine failures.

4.1.2. Data analysis

The data analysis section within our study is a critical component of the comprehensive data science lifecycle, encompassing a systematic series of interconnected steps aimed at distilling meaningful insights from raw datasets, as shown in Figure 3. Beginning with data preprocessing, we systematically addressed prevalent issues, including the treatment of missing values and outliers, employing robust techniques, including the Z-score and interquartile range (IQR) methods, which were utilized to detect and manage outliers through transformation or removal. Furthermore, to enhance computational efficiency, we applied dimensionality reduction techniques. The data required normalization due to variations in scale and units, achieved through min-max scaling to ensure uniform contributions of each feature.

Following the preparation phase, exploratory data analysis (EDA) facilitated a nuanced understanding of three distinct datasets, revealing distribution patterns and elucidating relationships among key parameters. EDA emerged as an invaluable guide for subsequent learning endeavors. EDA is indispensable for guiding data selection and ensuring the seamless execution of machine learning tasks [21]. In this research a variety of visualizations were employed to conduct EDA, enabling a comprehensive understanding of the three distinct datasets. These comprehensive preprocessing measures aimed to enhance the dataset's readiness for accurate and meaningful analysis and modeling.

The subsequent phase, feature engineering, concentrated on augmenting datasets to bolster pattern recognition within predictive models. Thoughtful generation of new features enriched the datasets, while strategic feature selection streamlined data, optimizing model performance [22]. Feature engineering assumes a central role in enhancing the datasets to empower the predictive models with improved pattern recognition capabilities [23]. As explained in our study, generating new features is a part of feature engineering that includes creating new variables, guided by either domain knowledge or mathematical transformations [24]. These new features enrich the dataset, potentially capturing previously hidden relationships and patterns. To streamline the dataset and enhance model efficiency, feature selection strategies come into play [25]. Various techniques were employed such as recursive feature elimination (RFE), information gain, correlation analysis, Chi-Square tests, and fisher score. These methods facilitated the identification and preservation of the most informative and influential features, contributing to dimensionality reduction and enhanced model performance. Finally, feature extraction aims to recognize the necessity to address high-dimensional data, we leverage principal component analysis (PCA) to restructure the dataset from a complex, multi-dimensional space into a more manageable, lower-dimensional representation. This process enhances computational efficiency while preserving essential data characteristics.

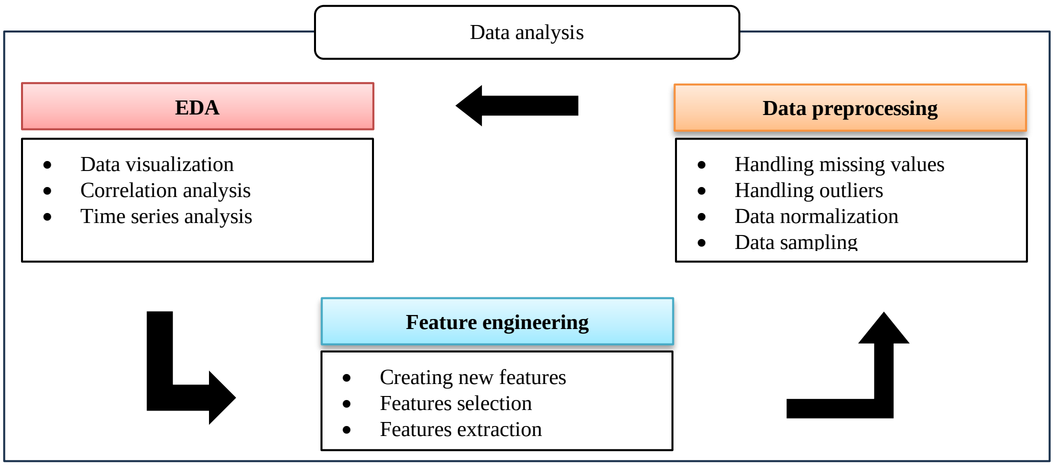


Figure 3. Data analysis and modeling workflow

4.2. Proposed architecture

Following our comprehensive evaluation of trained models, we proceeded to develop our innovative hybrid LSTM-SVM model. Figure 4 provides a visual representation of the architecture of this hybrid classifier, which consists of four pivotal layers: the input layer, hidden layer, fully connected (FC) layer, and softmax layer. Initially, 60% of the feature-extracted data from the count vectorizer is allocated for training, while 20% is designated for validation and another 20% for testing. We begin by feeding the training data into the LSTM input layer, initiating the flow of information. This input data proceeds into the hidden layer, a pivotal component within the LSTM model. The LSTM's hidden layer encompasses a complex mechanism with four interconnected layers that collaboratively produce the cell's output and state. Following this, both the output and state are passed on to the subsequent hidden layer. Distinguishing itself from conventional RNNs, LSTM introduces a more intricate structure. It incorporates not only a single tanh layer but also three logistic sigmoid gates. These gates determine which information to retain and which to discard as data traverses the network. Upon completion of the journey through the hidden layers, the output is directed to the FC layer. Situated higher in the network hierarchy, the FC layer plays a critical role in revealing the precise structures of features detected by the lower network layers. Here, the input is condensed into a dense feature representation, with each node within the FC layer independently learning its set of weights relative to all other nodes in the layer. Subsequently, the output from the FC layer is forwarded to the SoftMax layer, typically the final layer in LSTM networks. The softmax layer employs a SoftMax function, similar to the sigmoid function used in logistic regression, making it suitable for multi-class classification tasks. This is particularly valuable when classes are mutually exclusive. In our proposed model, we enhance the LSTM network's SoftMax layer by introducing SVM classifier. The SVM classifier is integral to both the training and classification phases.

During the training phase, a document is provided as input to the LSTM network, allowing the construction of a statistical model for the LSTM neural network. Subsequently, a feature vector is computed for each item in the training dataset, utilizing the weights obtained from the penultimate network layer. This training process is facilitated by the SVM classifier. In the classification phase, the same stages are replicated for each vector requiring classification. An embedding vector is derived using the previously trained LSTM network. Once the training process is completed, the trained model undergoes testing using the designated 20% of the feature-extracted data, while the remaining 20% serves as the validation set. Finally, the SVM classifier combines the embedding vector with other classification features to make precise class predictions.

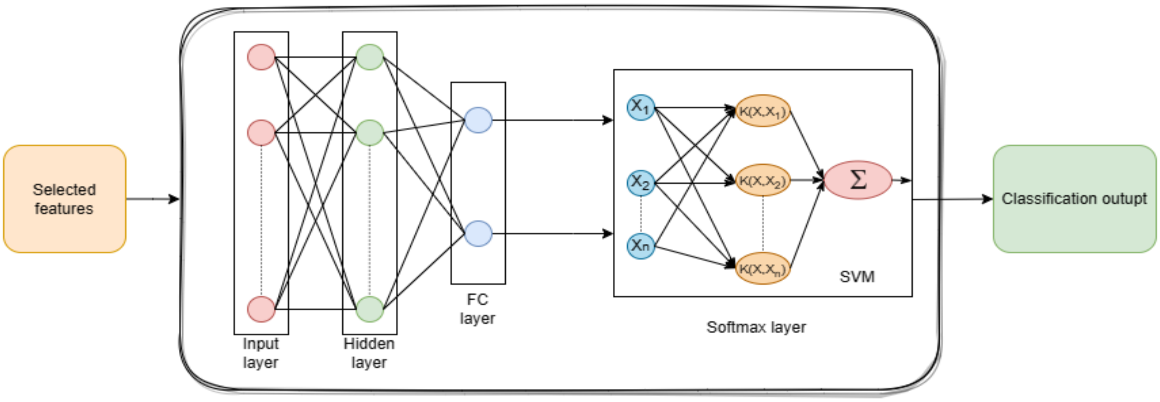


Figure 4. Hybrid model classifier LSTM-SVM architecture

4.3. Performance evaluation metrics

Various evaluation metrics are employed to assess the predictive performance of an experimental model's outputs. The evaluation criteria collectively provide a comprehensive understanding of the model's strengths. These metrics offer valuable insights into the model's effectiveness in classification tasks. In our study, we employ accuracy, precision, recall, and F1-score, a set of well-established metrics to rigorously evaluate the performance of our model. The essence of evaluation methods revolves around the identification of true negatives (TN), true positives (TP), false negatives (FN), and false positives (FP) in binary outcomes [26]. These evaluation metrics are shown in (1) to (4).

In this context, TN indicates the instances where the model correctly predicts non-failure (negative class), and indeed, there is no failure. On the other hand, TP represents the instances where the model correctly identifies failures, aligning with the actual occurrences of failure. However, FN represents the instance where the model incorrectly predicts non-failure when failure occurs, including instances where the model misses these failures. Conversely, FP represents the instances where the model incorrectly predicts failure, indicating cases where the model suggests failure, but no actual failure occurs [27].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1 - score = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{4}$$

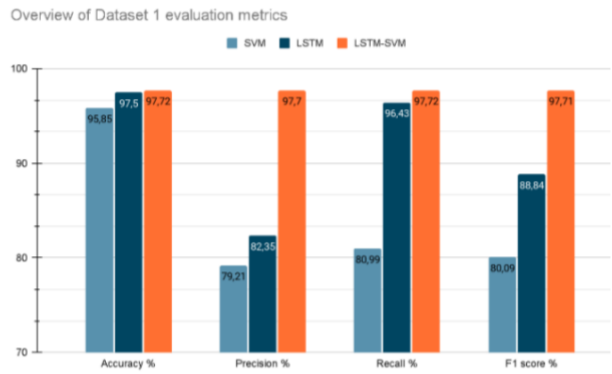
5. RESULTS AND DISCUSSION

In this section, we investigate the influence of different network architectures on model performance. We developed three distinct models for experimental comparison: SVM, LSTM, and a hybrid LSTM-SVM model. Our evaluation of classification results primarily centers around key metrics, with a primary emphasis on the F1-score, as detailed in this section. A comprehensive summary of all evaluation metrics is shown in Table 2.

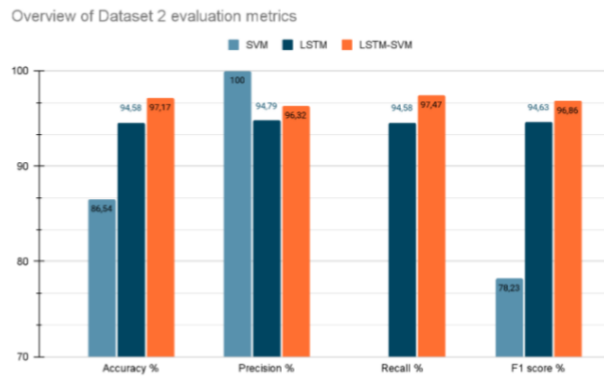
Table 2. Experimental results on different models

Data id	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Data 1	SVM	95.85	79.21	80.99	80.09
	LSTM	97.50	82.35	96.45	88.84
	LSTM-SVM	97.72	97.70	97.72	97.71
Data 2	SVM	86.54	100	64.25	78.23
	LSTM	94.58	94.79	94.58	94.63
	LSTM-SVM	97.17	96.32	97.47	96.86
Data 3	SVM	92.50	96.40	80.70	87.90
	LSTM	97.50	96.94	95.60	96.30
	LSTM-SVM	98.14	97.42	98.50	97.90

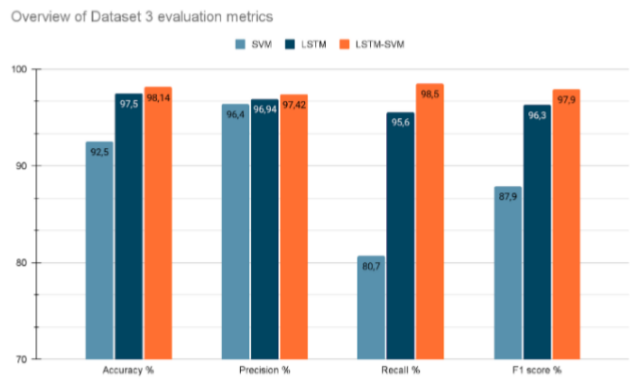
The three models produced different results. To better compare the prediction results, we show accuracy, precision, recall, and F1-score values in Figures 5(a) to 5(c) for these models across three datasets. These experiments were conducted based on the datasets introduced in section 4, providing a robust foundation for our analysis.



(a)



(b)



(c)

Figure 5. Overview of datasets evaluation metrics: (a) data 1, (b) data 2, and (c) data 3

When collecting data from IoT devices in natural environment scenarios, it is expected that the resulting datasets will exhibit class imbalances, as observed in the datasets we utilized. In such scenarios, if the classifier simply predicts each instance as belonging to the majority class and relies on overall classification accuracy for performance evaluation, it may yield artificially high accuracy scores. Consequently, using overall classification accuracy as the primary performance metric is inappropriate. The F-measure (F1-score) is a more suitable choice, as it considers both FP and FN, combining two measures known as 'precision' and 'recall' from the information retrieval community.

For instance, having 100% precision in data 2 for the SVM results means that the model rarely makes FP predictions, but this comes at the expense of recall, which is 64.25%. This indicates that the model correctly identifies a significant portion of actual positive cases (TP) but misses a substantial number of them (FN). While an accuracy of 86.54% may appear favorable, it can be significantly influenced by the class distribution. In cases where the cost of FN (missed failures) is considerable weight, accuracy alone is an insufficient metric. A model with a high F1-score is more suitable because it finds a good balance between precision and recall, recognizing both FP and FN. A higher F1-score indicates a better overall model performance, making it a more useful metric for imbalanced datasets or situations where there's a trade-off between FP and FN. The performance of the three models on three datasets based on the F1-score metric shown in Figure 6.

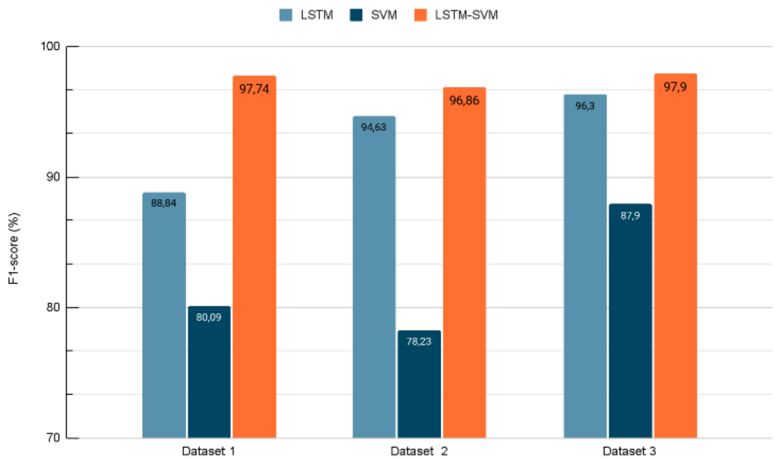


Figure 6. Performance of the three models on the three datasets

For dataset 1, our hybrid model achieved an impressive accuracy rate of 97.72%. Breaking down the individual models, the SVM model produced an F1-score of 80.09%, while the LSTM model yielded an F1-score of 88.84%. Notably, the hybrid LSTM-SVM model stood out with an outstanding F1-score of 97.71%. This is a significant improvement, showcasing an increase of about 8.87% over the LSTM model's performance. Turning our attention to dataset 2, the hybrid model attained an accuracy rate of 97.17%. In terms of F1-scores, the SVM model recorded 78.23%, while the LSTM model excelled with a score of 94.63%. Remarkably, the combination of models within the hybrid LSTM-SVM model delivered an F1-score of 96.86%. Here, the hybrid model showcased a substantial improvement of approximately 2.23% over the LSTM model's performance. Lastly, for dataset 3, the SVM model achieved an F1-score of 87.90%, while the LSTM model displayed its effectiveness with an F1-score of 96.30%. Surpassing both, the hybrid LSTM-SVM model demonstrated an impressive F1-score of 97.90%. In this instance, the hybrid model exhibited a notable increase of about 1.6% over the LSTM model's performance. These results consistently highlight the superiority of the hybrid LSTM-SVM model over individual SVM and LSTM models across different datasets. The hybrid approach proves to be a promising strategy for enhancing classification performance, particularly in the context of failure prediction tasks.

6. CONCLUSION AND FUTURE DIRECTIONS

Our innovative approach capitalizes on the strengths of both LSTM neural networks and SVM, yielding remarkable results in the domain of failure prediction. The LSTM component excels in capturing temporal patterns and dependencies within our industrial data, providing valuable insights into the evolving behavior of systems leading up to failures. Conversely, the SVM component efficiently classifies these patterns, prioritizing robust predictions, particularly in terms of the F1-score. Our study emphasizes the critical importance of predictive maintenance in industrial settings. By accurately identifying impending failures well in advance, we empower ourselves to proactively schedule maintenance interventions, effectively mitigating costly downtime and production disruptions. This, in turn, not only enhances operational efficiency but also translates into significant cost savings. In a world where data-driven decision-making reigns supreme, predictive maintenance, driven by advanced machine learning models, stands as a transformative force. It not only guarantees the reliability and sustainability of industrial operations but also lays the foundation for a more efficient and cost-effective future. Looking ahead, our future work envisions the continued refinement and enhancement of the hybrid model. We aim to leverage a variety of real-time datasets to optimize its performance further. Additionally, we plan to compare the LSTM-SVM hybrid model with other innovative hybrid models and explore its applicability in various fields, including emerging topics like green IoT. Through these endeavors, we aspire to advance the frontiers of machine learning applications and continue contributing to the advancement of predictive maintenance and industrial efficiency.

ACKNOWLEDGEMENTS

We express our sincere gratitude to YosoBox SARL Company and the University Mohammed First in Oujda, Morrocco, for their invaluable support and provision of essential facilities for this research.

REFERENCES

[1] M. F. Elrawy, A. I. Awad, and H. F. A. Hamed, "Intrusion detection systems for IoT-based smart environments: a survey," *Journal of Cloud Computing*, vol. 7, no. 1, 2018, doi: 10.1186/s13677-018-0123-6.

[2] A. R. H. Hussein, "Internet of things (IoT): research challenges and future applications," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, pp. 77–82, 2019, doi: 10.14569/ijacsa.2019.0100611.

[3] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014, doi: 10.1109/JIOT.2014.2306328.

[4] P. Nunes, J. Santos, and E. Rocha, "Challenges in predictive maintenance–a review," *CIRP Journal of Manufacturing Science and Technology*, vol. 40, pp. 53–67, 2023, doi: 10.1016/j.cirpj.2022.11.004.

[5] S. Namuduri, B. N. Narayanan, V. S. P. Davuluru, L. Burton, and S. Bhansali, "Review—deep learning methods for sensor based predictive maintenance and future perspectives for electrochemical sensors," *Journal of The Electrochemical Society*, vol. 167, no. 3, Feb. 2020, doi: 10.1149/1945-7111/ab67a8.

[6] G. A. Susto, A. Schirru, S. Pampuri, S. McLoone, and A. Beghi, "Machine learning for predictive maintenance: A multiple classifier approach," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 3, pp. 812–820, 2015, doi: 10.1109/TII.2014.2349359.

[7] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. D. P. Francisco, J. P. Basto, and S. G. S. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers and Industrial Engineering*, vol. 137, 2019, doi: 10.1016/j.cie.2019.106024.

[8] J. Leukel, J. González, and M. Riekert, "Adoption of machine learning technology for failure prediction in industrial maintenance: A systematic review," *Journal of Manufacturing Systems*, vol. 61, pp. 87–96, 2021, doi: 10.1016/j.jmsy.2021.08.012.

[9] M. S. Mahdavinejad, M. Rezvan, M. Barekatain, P. Adibi, P. Barnaghi, and A. P. Sheth, "Machine learning for internet of things data analysis: a survey," *Digital Communications and Networks*, vol. 4, no. 3, pp. 161–175, 2018, doi: 10.1016/j.dcan.2017.10.002.

[10] G. V. Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model ," *Artificial Intelligence Review*, vol. 53, pp. 5929–5955, 2020.

[11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[12] N. Sapankevych and R. Sankar, "Time series prediction using support vector machines: a survey," *IEEE Computational Intelligence Magazine*, vol. 4, no. 2, pp. 24–38, 2009, doi: 10.1109/MCI.2009.932254.

[13] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1023/A:1022627411411.

[14] J. Dalzochio *et al.*, "Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges," *Computers in Industry*, vol. 123, Dec. 2020, doi: 10.1016/j.compind.2020.103298.

[15] M. Yang and J. Zhang, "Data anomaly detection in the internet of things: A review of current trends and research challenges," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 9, pp. 1–10, 2023, doi: 10.14569/IJACSA.2023.0140901.

[16] A. Jain and A. M. Kumar, "Hybrid neural network models for hydrologic time series forecasting," *Applied Soft Computing Journal*, vol. 7, no. 2, pp. 585–592, 2007, doi: 10.1016/j.asoc.2006.03.002.

[17] A. Wahid, J. G. Breslin, and M. A. Intizar, "Prediction of machine failure in industry 4.0: A hybrid CNN-LSTM framework," *Applied Sciences*, vol. 12, no. 9, 2022, doi: 10.3390/app12094221.

[18] A. Borré, L. O. Seman, E. Camponogara, S. F. Stefenon, V. C. Mariani, and L. D. S. Coelho, "Machine fault detection using a hybrid CNN-LSTM attention-based model," *Sensors*, vol. 23, no. 9, May 2023, doi: 10.3390/s23094512.

[19] C. H. Yeh, M. H. Lin, C. H. Lin, C. E. Yu, and M. J. Chen, "Machine learning for long cycle maintenance prediction of wind turbine," *Sensors*, vol. 19, no. 7, 2019, doi: 10.3390/s19071671.

[20] K. Vos, Z. Peng, C. Jenkins, M. R. Shahriar, P. Borghesani, and W. Wang, "Vibration-based anomaly detection using LSTM/SVM approaches," *Mechanical Systems and Signal Processing*, vol. 169, 2022, doi: 10.1016/j.ymssp.2021.108752.

[21] K. Sahoo, A. K. Samal, J. Pramanik, and S. K. Pani, "Exploratory data analysis using python," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 12, pp. 4727–4735, Oct. 2019, doi: 10.35940/ijitee.L3591.1081219.

[22] I. Guyon and A. Elisseefl, "An introduction to feature extraction," in *Studies in Fuzziness and Soft Computing*, vol. 207, Berlin, Heidelberg: Springer, 2006, pp. 1–25. doi: 10.1007/978-3-540-35488-8_1.
[23] T. Verdonck, B. Baesens, M. Óskarsdóttir, and S. V. Broucke, "Special issue on feature engineering editorial," *Machine Learning*, Aug. 2021, doi: 10.1007/s10994-021-06042-2.
[24] P. Duboue, "Feature engineering: human-in-the-loop machine learning," in *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications*, Cham, Switzerland: Springer, 2022, pp. 109–127. doi: 10.1007/978-3-030-88389-8_7.
[25] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/j.compeleceng.2013.11.024.
[26] K. M. Ting, "Confusion matrix," in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 2011, pp. 209–209. doi: 10.1007/978-0-387-30164-8_157.
[27] Z. A. Bukhsh, A. Saeed, I. Stipanovic, and A. G. Doree, "Predictive maintenance using tree-based classification techniques: A case of railway switches," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 35–54, Apr. 2019, doi: 10.1016/j.trc.2019.02.001.

BIOGRAPHIES OF AUTHORS

Ouiam Khattach is a Ph.D. candidate in Computer Engineering at Mohammed First University in Oujda, Morocco, where she is researching internet of things failures using machine learning approaches. Graduated and received an engineering degree in big data and cloud computing from the school ENSET Mohammedia, Morocco in 2021. She participates in several scientific & organizing committees of national and international conferences. Additionally, she holds several certifications in artificial intelligence, programming, and networking. She can be contacted at email: ouiam.khattach@ump.ac.ma.

Prof. Dr. Omar Moussaoui is an associate professor at the Higher School of Technology (ESTO) of Mohammed First University, Oujda, Morocco. He has been a member of the Department of Computer Science at ESTO since 2013. He is currently the director of the MATSI Research Laboratory. Omar completed his Ph.D. in computer science at the University of Cergy-Pontoise France in 2006. His research interests lie in the fields of IoT, wireless networks, and security. He has actively collaborated with researchers in several other computer science disciplines. He participated in several scientific & organizing committees of national and international conferences. He served as a reviewer for numerous international journals. He has more than 20 publications in international journals and conferences and he has co-authored 2 book chapters. He is an instructor for CISCO networking academy on CCNA routing & switching and CCNA security. He can be contacted at email: o.moussaoui@ump.ac.ma.

Mohammed Hassine is a distinguished professional with over 20 years of expertise in product development, prominently contributing to the success of two startups, Raidtec Corporation and MPSTOR. As a key member of the management team at MPSTOR, he played a pivotal role in the company's triumphant exit, securing a multi million $ deal in 2016. With a focus on cyber security and AI, Mohammed's executive role at MPSTOR involved leading the development of intricate storage controllers, overseeing both hardware and software aspects. His responsibilities spanned high-level product design, strategic marketing initiatives, and the orchestration of all engineering processes necessary for transforming conceptual ideas into fully industrialized products. His project management expertise was instrumental in successfully delivering complex projects to Tier 1 customers, including industry giants such as EMC, Intel, Quantum, Avid, and AIC. His multifaceted role showcased a blend of technical proficiency, strategic vision, and leadership acumen, establishing Mohammed as a driving force behind the achievements of the organizations he has been a part of, particularly in the realms of cyber security and AI. He can be contacted at email: mo.hassine@tisalabs.com.