XAI-Driven Deep Learning for Real-Time Wireless Sensor Failure Prediction in Healthcare

Naima Samout
ISET Gafsa Tunisia
SETIT Lab
ENIG Tunisia

Thouraya Gouasmi
University of Gafsa,
ISSATG, Tunisia

Nejah Nasri
University of Gafsa,Tunisia
SETIT Laboratory Sfax, Tunisia

***Abstract*—Medical equipment predictive maintenance is essential to maintaining consistent and dependable healthcare services. Wireless Sensor Networks are essential for keeping an eye on medical devices and anticipating malfunctions before they happen. In this work, we provide a predictive maintenance strategy for medical WSNs based on LSTMs and use Local Interpretable Model-Agnostic Explanations to improve its interpretability. Our method increases the accuracy of fault predictions while providing decision-making transparency. Results from experiments show how well our model works in real-time to explain contributing elements and spot possible failures..**

I. INTRODUCTION

The real-time monitoring of critical metrics and medical devices, especially treatment-related pumps, has advanced significantly as a result of the Internet of Things (IoT) in healthcare systems [1], [2]. However, to guarantee appropriate operation and reduce patient safety hazards, accurate forecasts regarding device failures are crucial to the efficient and dependable management of these systems. In medical IoT equipment, particularly pumps, prompt failure detection is essential to guarantee patient safety and avoid serious repercussions in the event of a malfunction. Implementing machine learning models that can identify anomalies and anticipate errors before they happen is crucial given the enormous volume of data produced by these sensors. Even though they are good at detecting anomalies, complicated predictive models like Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks are frequently referred to as "black boxes," which makes it challenging to comprehend the logic underlying the model's judgments [3]. This lack of openness can be problematic, particularly in delicate domains like healthcare, where automated systems' decisions need to be comprehensible and justified. Using an LSTM model for anomaly detection, this study aims to provide a real-time failure prediction method for pumps in a Wireless Sensor Network (WSN) for medical applications [4]–[6]. In order to maintain a high degree of accuracy and dependability in the forecasts, this study uses model explainability techniques, such as LIME (Local Interpretable Model-agnostic Explanations), to make the prediction results easier for end users to understand [7].

The main controbutions of this paper are:

- Development of a real-time failure prediction model: Using an LSTM model to analyze data from IoT sensors, enabling accurate prediction of pump failures in healthcare settings.
- Improvement of prediction explainability: Integrating the LIME technique to explain the predictions made by machine learning models, making decisions more interpretable and justifiable, especially for healthcare professionals.
- Implementation of a practical solution for medical IoT systems: Developing a real-time system capable of analyzing sensor data, predicting failures, and providing understandable explanations to users, taking into account the constraints of IoT systems deployed in medical environments.

This work is organized as follows: the majority of Section 2 discusses the background and related works. Section 3 presents the Methodology of our paper, and Section 4 provides provides experimental results and a discussion on the findings. Finally, Section 5 concludes the paper

II. RELATEK WORK

Machine learning approaches for predictive maintenance have been investigated in a number of research. Traditional techniques like Random Forests and Support Vector Machines (SVM) have been employed, however they are ineffective at capturing sequential dependencies. In time-series prediction, deep learning models like as Transformers and LSTMs have demonstrated encouraging outcomes. These models are not interpretable, though. Recent improvements in Explainable AI approaches, such as SHAP and LIME [8], [9], enable understanding of model predictions. For interpretable predictive maintenance in healthcare WSNs, our work integrates LSTM and LIME.

The study [10] offers a thorough framework for deploying explainable and transparent artificial intelligence sensors in the healthcare industry in order to overcome the difficulties presented by AI "black boxes" and comply with the requirements of the European Union's (EU) AI Act and Data Act,. To guarantee that AI sensor outputs are understandable, auditable, and consistent with clinical decision-making, our methodology integrates interpretable machine learning (ML), human–AI interaction, and ethical standards.

In the context of semisupervised FM recognition and RUL prediction, this research [11] suggests a deep learning network

for adaptive sensor selection, specificallyin situations where there are several OCs. The excellence, universality, and scalability of the suggested approach are shown by the outcomes of these thorough assessments and equitable comparisons with cutting-edge techniques. In particular, the suggested approach extracts OC-invariant features that remove domain disparities resulting from different OCs and accomplishes adaptive and relevant sensor selection customized to distinct FMs. Additionally, it shows minor RUL prediction errors along with excellent FM recognition accuracy when compared to current FM recognition and RUL prediction approaches.

This paper [12] proposes a deep-learning network designed for adaptive sensor selection in the context of semisupervised FM recognition and RUL prediction. tey prouved that this deep learning network can be applied effectively to various degraded machines experiencing multiple FMs and OCs. It is particularly suitable for machines with complex physical mechanisms or unknown failure thresholds.They found that the approach is a good tool for getting insights and estimating failure patterns and frequency of faults, which our research tries to provide.

The authors [13] proposed a Fault Prediction in Wireless Sensor Networks using Soft Computing. Using ARIMA, they predicted failures by decomposing time series data into trend, seasonality and residual components and fit a model.As a result, they demanstrated the robustness forecast estimation model with confidence bounds: 80% and 95% is developed which can help in implementing an intelligent IoT infrastructure in WSN.

The authors [14] discussed the usefulness of the Explainable AI (EXAI) and its applications. Opportunities and possible project implementations in the healthcare sector, as well as the field of related applications like ECG, are also covered in the conversation. They discussed the influence of explainable AI on AI jargon and its wide range of uses in the healthcare sector, which were not covered in the previous surveys with all the features and characteristics. A solution taxonomy for EXAI-based medical-assisted programmable techniques has been offered by the authors after they have evaluated the most recent literature. The design, architecture, healthcare ecosystems, and end-to-end communication explanations of EXAI that underpin the fundamental ideas of DL are also presented in the study.

This paper [15] has covered the principles, implementation, and performance of a number of well-known XAI techniques in medical picture applications have all been . Initially, the different algorithms were divided into several different groups. A well-known XAI pertaining to medical picture classifications was examined in the context of AI currently used in medical imaging domains. A synopsis of how the recently suggested XAI techniques were used to improve the interpretability of their suggested models was also provided. Additionally, the necessity of explainable models for radiomic analysis was examined and clarified. A summary of this section is given in Table I, which summarizes the advantages and disadvantages of the current detection methods compared to our proposed model.

III. Proposed Methodology

This work introduces an Explainable AI-driven Long Short-Term Memory (XAI-LSTM) model-based real-time anomaly detection framework for healthcare Wireless Sensor Networks that combines real-time data collection, preprocessing, model training, and interpretability through the use of Local Interpretable Model-Agnostic Explanations (LIME). the work fowchart is presented in fig. 1).

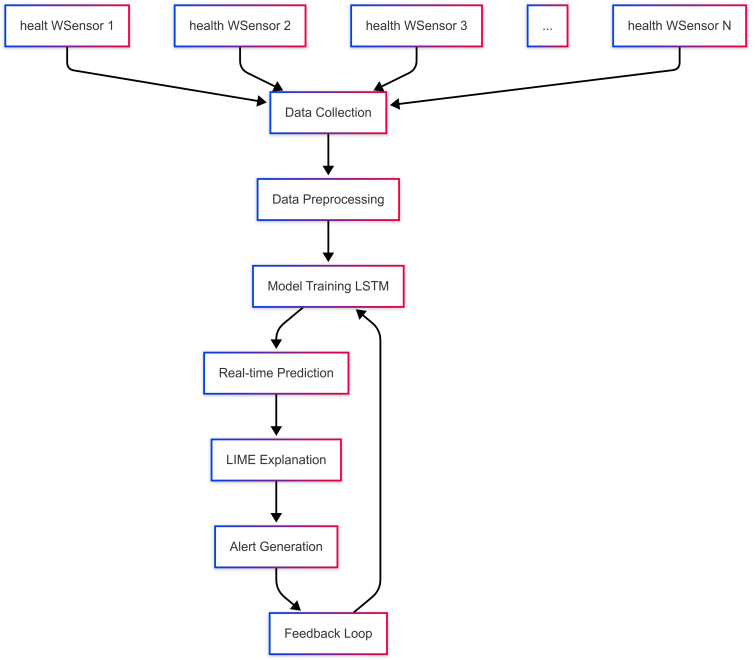

Fig. 1. Proposed framework of XAI-LSTM.

A. Data Collection

Real-time WSNs set up in a medical setting provided the data for this study. These sensors are crucial for patient health monitoring since they continuously measure a number of characteristics, including temperature, pressure, and heart rate. The dataset comprises labels for both normal and failure conditions along with a time series of sensor measurements. To ensure prompt failure detection, each sensor produces data at regular intervals, usually once every minute.

B. Data Preprocessing

Preprocessing the data is essential to guaranteeing the model's input quality. To preserve data integrity, any missing, inaccurate, or outlier values in the dataset are first treated via imputation or removal. Normalization is used to scale the input features into a consistent range because sensor data might vary greatly between devices and time periods. This ensures that every feature contributes equally to the model. In order to capture temporal dependencies—which are crucial for LSTM models—feature engineering is then carried out by developing

TABLE I
COMPARISON OF DIFFERENT APPROACHES IN AI-DRIVEN PREDICTION AND EXPLAINABILITY

Reference	Goal(s)	Novel Approach	Performance Metric	Dataset	Results	Strengths	Weaknesses	Explainable AI
[10]	Compliant AI for Healthcare WSNs	Interpretable ML, Human-AI Interaction	Interpretability, Auditable AI	Healthcare sensor datasets	Compliance with EU AI Act	Ensures transparent AI decisions	May sacrifice accuracy for explainability	XAI
[11]	Adaptive sensor selection for fault prediction	Adaptive Deep Learning Model	FM Recognition, RUL Prediction	Sensor failure datasets	High FM recognition accuracy	Reduces domain discrepancies	No explicit explainability	-
[12]	Understanding failure patterns	Deep Learning for failure prediction	Failure estimation accuracy	Various degraded machine datasets	Effective for unknown failure thresholds	Good insights into failure patterns	Opaque model behavior	-
[13]	Time-series forecasting for failure prediction	ARIMA (Time Series)	Prediction Accuracy	WSN datasets	80%-95% confidence bounds	Strong theoretical basis	Limited to linear patterns	XAI
[14]	Explainable AI in Healthcare	EXAI for Medical AI	Transparency in AI decisions	Healthcare AI datasets	Improved AI adoption in clinical settings	AI governance & explainability focus	Computational complexity	XAI
[15]	XAI for Medical Imaging	Deep Learning + XAI	Medical Image Classification Accuracy	Radiomics & Medical Imaging datasets	High accuracy in medical AI	Enhances interpretability for clinicians	-	XAI

time-based features including rolling averages, statistical measures (mean, variance), and windowing approaches. In order to allow the model to learn to differentiate between normal and defective sensor circumstances, the dataset is finally classified into two categories: "Normal" and "Failure." Failure events are associated with anomalous sensor activity that may point to possible failures.

C. LSTM Model

In order to predict sensor failures, the LSTM network is selected due to its capacity to identify temporal relationships in time-series data. The architecture is made up of multiple important layers: The preprocessed sensor data is supplied to the input layer in a sequential pattern, with each time step denoting a sensor reading. One or more LSTM units that identify temporal patterns in the data make up the LSTM Layer(s). These layers enable the model to understand the time-based correlations between sensor readings by storing data from earlier time steps.In order to convert the LSTM output into the final prediction and ascertain if the sensor is operating normally or exhibiting early indications of failure, a Dense Layer is added after the LSTM layers. The model determines the sensor's condition by using a single neuron in the final output layer with a sigmoid activation function to classify the sensor's status as either "Normal" or "Failure".

D. Model Training

A time-series classification method is used to train the LSTM model in order to forecast sensor failures. A training-validation split, in which the dataset is normally split into 80% for training and 20% for validation, is the first step in the training process. To help direct the learning process for classification tasks, the model measures the discrepancy between the actual labels and the projected probabilities using the Binary Cross-Entropy Loss Function.The Adam optimizer, a gradient descent variation that modifies the learning rate during training to promote effective convergence and enhance model performance, is used for optimization. Early Stopping, which ensures that the model generalizes effectively to unseen data, is applied based on validation loss to prevent overfitting. If the validation loss does not improve over a predetermined period of epochs, the training process stops.

E. LIME for Model Interpretability

LIME analyzes feature importance and perturbs input samples to produce locally interpretable approximations of the LSTM predictions. This enables medical practitioners to identify the sensor values that may be linked to malfunctions. Given that deep learning models—especially LSTMs—are sometimes seen as "black boxes", interpretability is crucial in order to comprehend the reasoning behind the prediction of a sensor failure. The model's decision-making process is transparent because to the use of Local Interpretable Model-agnostic Explanations (LIME). LIME approximates the intricate, black-box model using more straightforward, interpretable models (like linear regression) that are locally true to the original model's predictions. LIME creates local surrogate models that resemble the behavior of the LSTM model in the vicinity of each prediction produced by the LSTM model by perturbing the input data. By highlighting the most significant elements that led to the anticipated failure, LIME provides insight into the logic of the model and guarantees that the forecasts are comprehensible and justified.

F. Model Evaluation Metrics

The binary classification problem frequently suffers from a very unbalanced distribution of classes in the context of predicting failure for WSNs in the healthcare industry. With less than 1% of the data falling into the failure class ("1") and more than 99% falling into the normal class ("0"), sensor failure events are uncommon in this scenario. For machine learning algorithms and conventional assessment criteria, this discrepancy poses a serious problem.Because of this, even though overall accuracy (OA) is frequently employed in classification tasks, it is not a suitable statistic for model evaluation. Accuracy is a problem since a model can score highly just by correctly predicting the majority class (normal operation) for the majority of data while neglecting the minority class (sensor failure), which is the prediction's main focus. As a result, sensor failure detection which is crucial in healthcare settings

performs poorly. Because they offer a more impartial assessment of the model's performance on both classes, precision, recall, and the F1 score are thus the recommended evaluation metrics. Therefore, the optimization loop for the model will focus on maximizing the metrics that best meet the healthcare organization's objectives and operational constraints, ensuring the most effective and reliable sensor failure prediction in real-time WSNs.

IV. Experiments and Results

The findings of the XAI-LSTM model's use of real-time data to forecast sensor failures in healthcare settings are shown in this section. Precision, recall, F1 score, Cohen's Kappa, and Area Under the ROC Curve (AUC-ROC) are used to assess the model's performance. The dataset used for the evaluation is extremely unbalanced, with most sensor readings being categorized as "Normal" (class 0) and sensor failures being considered an uncommon occurrence (class 1). Local Interpretable Model-agnostic Explanations (LIME) are used to evaluate the interpretability of the LSTM model in addition to its performance.

A. Model Performance

A dataset of sensor readings from a healthcare setting, where the objective is to forecast if a sensor would fail in a future time interval, was used to train and assess the XAI-LSTM model. Table II below provides a summary of the evaluation metrics' findings. Exceptionally high accuracy, precision, recall, and F1 score were all attained by our model.

TABLE II
Comparison of the Experimental Results

Model	Accuracy	Precision	F1-Score	Recall
Random Forest	99.78	99.43	99.43	99.43
XGBoost	97.71	97.5	97.58	97.58
DNN	98.3	98.8	98.8	98.7
Our Model	99.98	99.86	99.86	99.86

B. Model Interpretability with LIME

LIME charts offer a concise, visual description of each feature's influence on decision-making, both positive and negative, providing more in-depth explanations for why a specific choice was produced by the suggested model. The LIME plot for the top eight most significant features is shown in fig 2, illustrating how they affect the determination of whether the network is secure or under assault at a given time. A feature's positive impact on the decision-making process is indicated by a blue, horizontal bar that runs to the right in the plot. Conversely, a negative influence is indicated by a orange bar that extends to the left.

We found important characteristics that contribute to failure predictions using LIME. This makes it possible for maintenance crews to respond proactively before serious malfunctions happen. The outputs of the LIME is shown in fig 2.

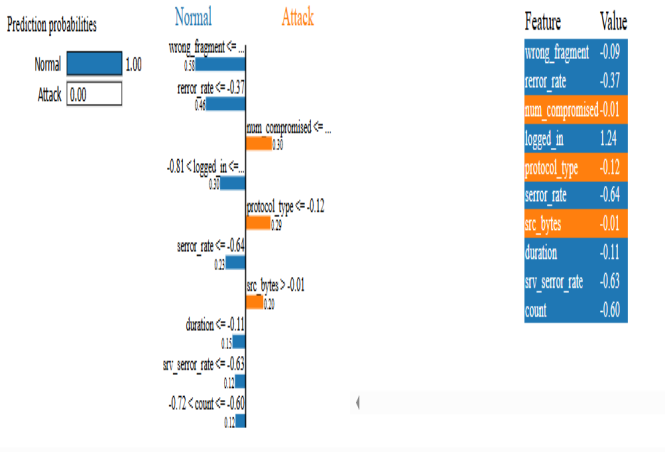

Fig. 2. XAI-based Interpretation of Model Predictions and Feature Contributions.

V. Conclusions

With the help of Explainable Artificial Intelligence (XAI) techniques, specifically LIME, we improve the interpretability of the model's predictions and provide important insights into how sensor features impact failure outcomes. This paper introduces a thorough method for deep learning-driven real-time failure prediction in wireless sensor networks within healthcare systems using LSTM models. The LSTM-based model outperforms conventional machine learning techniques like Random Forest, XGBoost, and Deep Neural Networks in terms of accuracy and reliability for real-time prediction tasks, and our results show that certain sensor readings greatly help distinguish normal operations from failures. Explainability using XAI not only promotes decision-making transparency but also increases user confidence in the system's results, which is crucial for vital healthcare applications. In future research, we aim to address two key challenges to further enhance the effectiveness and applicability of our predictive framework. Initially, we will investigate methods for enhancing model performance in situations with little labelled data, like transfer learning, data augmentation, or semi-supervised learning—all of which are frequent limitations in actual healthcare implementations. Second, we will look into ways to deal with domain shift issues brought on by discrepancies between the data used to train the model and the data that is encountered in real-time operation. In particular, we intend to use advanced domain adaptation strategies to make sure the model is resilient and performs well in a variety of changing healthcare contexts.

References

[1] Rejeb, A., Rejeb, K., Treiblmaier, H., Appolloni, A., Alghamdi, S., Alhasawi, Y., Iranmanesh, M.: The Internet of Things (IoT) in healthcare: Taking stock and moving forward. Internet of Things 22, 100721 (2023).

[2] S. Mnasri, N. Nasri, and T. Val, 'The deployment in the wireless sensor networks: Methodologies, recent works and applications', in *Proc. Int. Conf. Performance Evaluation and Modeling in Wired and Wireless Networks (PEMWN)*, 2014,

[3] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," Physica D: Nonlinear Phenomena, vol. 404, p. 132306, 2020.
[4] Islam, M.N.U., Fahmin, A., Hossain, M.S., Atiquzzaman, M.: Denial-of-service attacks on wireless sensor network and defense techniques. Wireless Personal Communications 116, 1993–2021 (2021).
[5] Ehteram, M., Afshari Nia, M., Panahi, F., Farrokhi, A.: Read-First LSTM model: A new variant of long short term memory neural network for predicting solar radiation data. Energy Conversion and Management 305, 118267 (2024).
[6] X. Feng, X. Ding, and S. Sun, "A security detection scheme based on evidence nodes in wireless sensor networks," in Proc. 6th Int. Conf. Biomed. Eng. Informat., Dec. 2013, pp. 689–693.
[7] Vimbi, V., Shaffi, N., Mahmud, M.: Interpreting artificial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection. Brain Informatics 11, 10 (2024).
[8] Wang, Y., Wang, A., Wang, D., Wang, D.: Deep Learning-Based Sensor Selection for Failure Mode Recognition and Prognostics Under Time-Varying Operating Conditions. IEEE Transactions on Automation Science and Engineering (2024).
[9] A.B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, et al. "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI." Information Fusion, Vol. 58, 2020, pp. 82-115. DOI: 10.1016/j.inffus.2019.12.012
[10] Bouderhem, R.: A Comprehensive Framework for Transparent and Explainable AI Sensors in Healthcare. Engineering Proceedings 82(1), 49 (2024). https://doi.org/10.3390/ecsa-11-20524
[11] Angelov, Plamen P and Soares, Eduardo A and Jiang, Richard and Arnold, Nicholas I and Atkinson, Peter M "Explainable artificial intelligence: an analytical review", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, (2021),
[12] Wang, Y., Wang, A., Wang, D., Wang, D.: Deep Learning-Based Sensor Selection for Failure Mode Recognition and Prognostics Under Time-Varying Operating Conditions. IEEE Transactions on Automation Science and Engineering (2024).
[13] Ara, T., Prabhakar, M., Bali, M.: Fault Prediction in Wireless Sensor Networks using Soft Computing. In: Proc. 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), pp. 532–538 (2020). https://doi.org/10.1109/ICSTCEE49637.2020.9277216
[14] Saraswat, D., Bhattacharya, P., Verma, A., Prasad, V.K., Tanwar, S., Sharma, G., Bokoro, P.N., Sharma, R.: Explainable AI for Healthcare 5.0: Opportunities and Challenges. IEEE Access 10, 84486–84517 (2022). https://doi.org/10.1109/ACCESS.2022.3197671
[15] Chaddad, A., Peng, J., Xu, J., Bouridane, A.: Survey of Explainable AI Techniques in Healthcare. Sensors 23(2), 634 (2023). https://doi.org/10.3390/s23020634