Team 7: AI-Enabled Anomaly Detection for Enhancing Meteorological Mast Sensor Data Quality

Vatsal Jitendra Patel, Latika Liladhar Dekate, and Lakshminarayanaa Rajamannar
Emails: vpate160@asu.edu, ldekate@asu.edu, lrajaman@asu.edu

***Abstract*— Modern Wind energy systems rely heavily on the quality of their operational data for efficient and reliable operation and predictive maintenance. The sensor data quality control (QC), often relies on static, rule-based algorithms that fail to address real-world scenarios and when executed by traditional methods, it often relies around static, rule-based algorithms which failed to address real world scenarios and failed to detect complex or novel anomalies. This paper proposes deep learning-based approach usi Long-Short Term Memory (LSTM) Autoencoder, that has enhanced anamoly detection in multivariate time-series from meteorological mast sensors. The approach we took is to design a model that learns the normal and optimal operational patterns of a system and identify anomalies by flagging data points with high reconstruction errors. The report analyses the LSTM Autoencoder for its robustness, dynamic and scalable characteristics by comparing it with the traditional rule-based QC script. This report pitches LSTM as a robust model that ensures data integrity for improving predictive and efficient O&M of Meteorological masts.**

***Keywords*—deep learning, anomaly detection, LSTM, autoencoder, predictive maintenance, time series, data quality control**

I. INTRODUCTION

The renewable energy shift is being critical for sustainability and carbon neutrality. Green generation is a prime resource provided by wind, universally accessible across the globe. This transition is highly dependent on the effective functioning of Meteorological mast Generators. Behind this transition is a massive sensor data acquisition of wind parameters such as speed, direction, pressure, and temperature in varying conditions. Such data may be subject to fault and inconsistencies because of sensor failure or harsh environments.

The data quality control is the main challenge. They're static and do not discover sophisticated sensor time-series data anomalies. Simple errors or not, they do not manage non-linear multivariate data relationships well. Missed fault discoveries in the early stages, inaccurate forecasts, and erroneous analysis result from undetected anomalies, both risking equipment failure or costly downtime.

This paper introduces an application for data quality control using an LSTM Autoencoder that was motivated by deep learning's performance in anomaly detection and proceeds to model normal sensor behavior and raise flags on departures, improving data validation.

II. COMPARISON OF QUALITY CONTROL METHODOLOGIES

A. Traditional Rule-Based Method

The typical method of controlling for data quality, employed in the simple script for the project, is that of a stepwise system with fixed checks. This is a system that relies on predetermined rules and fixed statistical limits in checking for data. Key checks in such a system:

- **Range Checks:** Ensuring data points fall within a plausible, physically possible range (e.g., wind speed cannot be negative).
- **Step Checks**: Flagging improbable jumps between consecutive readings.
- **Internal Consistency Checks:** Verifying that relationships between different parameters are logical (e.g., a zero-wind speed should not have a defined wind direction).

The traditional method is efficient in catching gross errors, this method is inherently limited. It is static and brittle; the rules must be manually defined and do not adapt to changing conditions like the uniqueness of terrain, varying weather, or different meteorological mast models without significant reprogramming. Most importantly, it cannot find complex problems that have many different factors. Each sensor reading may seem normal, but are very unusual.

B. AI-Driven Approach

The deep learning model in its present state is transitioning from hard and fast rules to dynamic learning. It uses an LSTM Autoencoder for learning regular working states from past data. Its benefits are stated below:

- **Adaptive and Resilient:** The model of artificial intelligence distinguishes errors by recognizing deviations from a learned model of normalcy, and not from the use of a fixed list of rules. This allows it to identify new kinds of errors.
- **Handles Complexity**: It is highly competent in comprehending complex, non-obvious, and time-related relationships among various sensors. This assists it in identifying small issues that would totally be ignored by a rule-based system.

- **Scalability:** After it is trained, the model can process information from multiple sources with little that needs to be changed, and is therefore more scalable than hand-coded programs. Text notes that deep learning models have been observed to achieve higher performance than traditional machine learning methods like Support Vector Machines and Random Forests in difficult fault diagnosis tasks.

III. LSTM Autoencoder for Finding Unusual Patterns

The detection method uses an LSTM Autoencoder, an artificial neural network autoencoder that is applicable for detecting nonstandard sequence data in an unsupervised setting.

A. Significant Building Blocks

The model consists of two primary components:

- **Autoencoder:** An unsupervised neural network that includes an encoder for compressing the data in the latent space and a decoder for reconstructing it.
- **LSTM (Long Short-Term Memory):** It is a special Recurrent Neural Network (RNN). LSTM cells are used in both the encoder and decoder because they can understand time-related patterns in data that comes in a sequence, which makes them perfect for this job.

B. Anomaly Detection Mechanism

The number one idea in the approach is to educate the model in the context of what is "normal." Reconstruction error motivates the whole process:

- **Training:** It is only trained on one set of clean sensor readings that contain no errors and have been through typical quality screening. The model's intention is to minimize the discrepancy between its original input and the output that it generates, such that it gains an in-depth understanding of regular system behavior.
- **Detection:** Once it is trained, the model is used on new data it has never seen before. When given an anomalous data point, it will have difficulty reconstructing it appropriately because the data is outside of the regular patterns it was trained on. This leaves it with having a significantly larger error in reconstructing it. Thresholding: With the setting of the limit for the reconstruction error, any data point that is beyond this limit is designated an anomaly.

IV. IMPLEMENTATION AND SIMULATED RESULTS

A. Data Preparation

The first step was to load the Wind_SampleData.csv dataset. To create a reliable training set, the traditional rule-based QC script was applied to the raw data. These points that had been marked as suspicious or erroneous were all discarded, leaving a pristine dataset of the typical operating state of the meteorological mast. Data was normalized using a MinMaxScaler such that all the features would all be in the 0 to 1 range, which is standard when it comes to training neural networks. Finally, the data was transformed into overlapping time-series sequences of 24 steps each to serve as input for the LSTM layers.

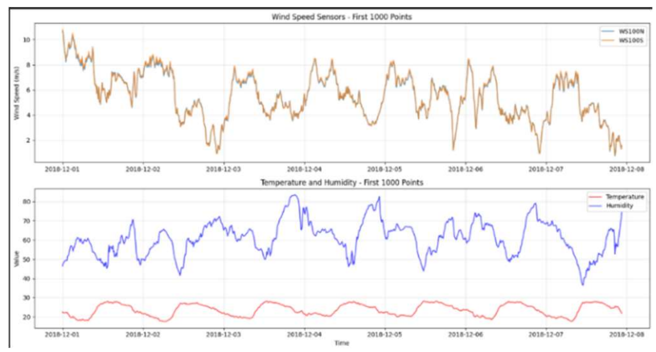

Figure 1: Raw Timeseries Dataset from Wind Speed Sensors before Quality Control measures applied.

B. Model Training and Thresholding

The LSTM Autoencoder was implemented using TensorFlow/Keras, and it was trained on clean data with Mean Absolute Error loss function and Adam optimiser. From the loss curves, the model is able to reconstruct the normal data, and the loss is falling and settling over epochs.

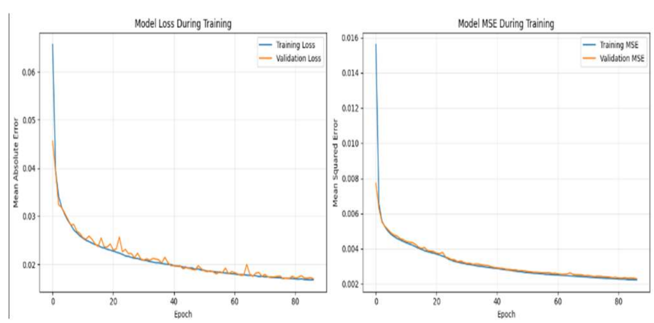

Figure 2: Model Training and Validation Loss

Figure 2 shows how the model learned to reconstruct regular data. After training, the model measured the reconstruction errors on the training data. The errors were graphed, and three times the standard deviations were added to the mean to set a threshold. This threshold created a line that separates normal behavior from unusual behavior.

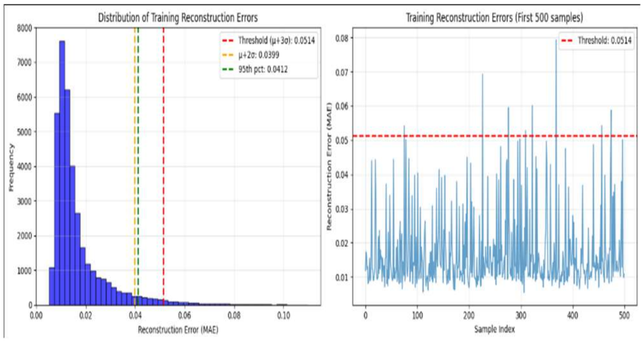

Figure 3: Error Distribution in the Training Data

C. Anomaly Detection Results

To put the model to the test, it was applied to the entire data set that contained both regular and anomalous points. The model computed the reconstruction error for every time step. Any point that had an error that was larger than the limit was flagged as an anomaly.

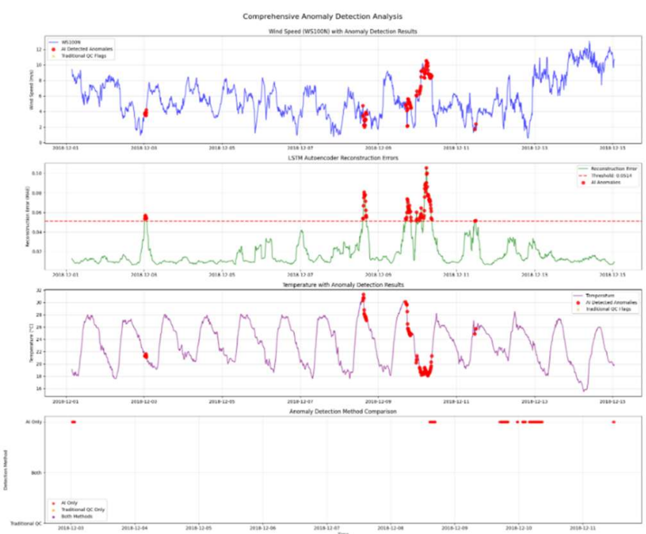

Figure 4: Proof-of-concept Visualization of Anomaly Detection

Figure 4 illustrates the output of the model, which shows points that do not follow normal trends. Points that have been marked for being anomalous by LSTM Autoencoder because they had a reconstruction error threshold violation, are represented with red dots.

IV. DISCUSSION AND CONCLUSION

A successful design was made for an AI application for anomaly detection in sensor data from meteorological masts. From the comparison of the rule-based approach and LSTM Autoencoder, it is revealed that deep learning is superior in terms of adaptability, complexity, and robustness. The AI model provides an in-line data quality control tool through learning of normal operation trends, in contrast to fixed-rule systems.

The proof-of-concept demonstrates that an LSTM Autoencoder is able to detect anomalies that have been missed in the past by established methods, improving predictive maintenance and reliability of meteorological mast Generators and investigating meteorological metmast data for Wind Farm viability studies. This report presents its design and outcomes, which support established research, validating the potential of its method.

Also, future work could involve investigating higher-level hybrid structures, for instance, the combination of an autoencoder and a Graph Attention Network. This could potentially enable learning of temporal patterns and learning of dynamic sensor relations for improved representation and, hence, anomaly detection.

REFERENCES

[1] B. Zhang and Q. Wu, "Simulation of Predictive Maintenance Model of Power Enterprise Equipment based on Deep Learning Algorithm," *2024 3rd International Conference on Energy and Power Engineering, Control Engineering (EPECE)*, Chengdu, China, 2024, pp. 1-4

[2] G. Zhu, L. Huang, D. Li and L. Gong, "Anomaly Detection for Multivariate Times Series Data of Aero-engine Based on Deep LSTM Autoencoder," *2024 6th International Conference on Electronics and Communication, Network and Computer Technology (ECNCT)*, Guangzhou, China, 2024, pp. 190-194

[3] J. Wang, "Enhancing Anomaly Detection in Multivariate Time Series with Hybrid Deep Learning Architecture," *2025 5th International Conference on Neural Networks, Information and Communication Engineering (NNICE)*, Guangzhou, China, 2025, pp. 1040-1045.

[4] Y. Chai and J. Hou, "Research on Fault Diagnosis and Predictive Maintenance of Electrical Automation System Based on Deep Learning," *2025 International Conference on Electrical Drives, Power Electronics & Engineering (EDPEE)*, Athens, Greece, 2025, pp. 1198-1202

[5] T. Sandro *et al.*, "Deep Learning Approach Using K-Nearest Neighbors and Neural Network for Quality Assurance of Water Temperature Sensor in Automatic Weather Station," *2023 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, Malang, Indonesia, 2023, pp. 147-152

APPENDIX A

GitHub: projectCode
Presentation Video Link: presentation