IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13365

ANOMALY DETECTION IN TIME SERIES DATA IN IoT ENVIRONMENT

Shibzan Shahanas[1], Afnaj Akthar[2], Saanna Anand[3], Rakshitha[4], Dr. Amirthavalli.M[5]

Student, Dept. of Artificial Intelligence & Machine Learning, Mangalore Institute of Technology & Engineering, Moodabidri, India[1-4]

Professor, Dept. of Artificial Intelligence & Machine Learning, Mangalore Institute of Technology & Engineering, Moodabidri, India[5]

Abstract: This project is about technique or approach in finding anomalies, which represents deviations from expected patterns, can signify critical events of irregularities, malfunctioning of sensors, demanding accurate detection. Internet Of Things (IoT) represents a framework that links physical devices to the internet, allowing them to communicate and exchange data. The quality of IoT services usually depends on the integrity and accuracy of the data. Time series is a common type of data found in everyday situations like traffic flow, network performances, financial records, etc. Detecting anomalies in time series IoT sensor data is very much needed because of the possibility of noise and unavailability of labels in the sensor readings and it's also an important research topic with practical uses such as spotting intrusions in networks, monitoring traffic, and identifying errors in sensor data. In this project the Inter-Berkeley Research Lab dataset is used for unlabeled anomaly detection technique and UNSW-NB15 IoT weather board sensor dataset is used for labelled anomaly detection, which is suitable for testing and validating different anomaly detection methodologies. This project is proposed to work on hybrid models such as , LSTM – Autoencoder +Isolation Forest, Bi – LSTM + OneClass SVM, an Ensemble model of DBSCAN, LOF, SVM, and a Statistical approach for anomaly detection in IoT sensor Time Series Data, using the results to understand better about the performance of these proposed models.

Keywords: Internet Of Things (IoT),Bidirectional Long Short-Term Memory(Bi-LSTM),One-Class Support Vector Machine(One-Class SVM),Density-Based Spatial Clustering of Applications with Noise.(DBSCAN),Local Outlier Factor (LOF).

I. INTRODUCTION

The Internet of Things have changed the way we live by attaching day-to-day things to the web. IoT devices gathers information from sensing units and also using it to make jobs less complicates. This exchange of information allows devices to interact as well as make educated decisions. It helps us to comprehend our actions, practices, and choices resulting in the tailored experiences. For example, IoT can recommend items or offer health and wellness suggestions based upon our information. Data – driven choices additionally boost effectiveness. IoT also boosts security as well as protection by discovering abnormalities or threats such as uncommon patterns in protection or surveillance systems.

Time series data, which tracks modifications in time is important for recognizing IoT sensing unit analyses. This information's time-based nature gives understandings right into patterns which helps us to find the irregularities. Irregularities are abnormalities that might suggest problems in IoT systems. Different techniques consisting of analytical plus deep learning- based strategies are utilized for discovering anomalies.

Many deep learning techniques are known for its ability to capture complex patterns over time, and so is widely used. But there are also some common problems occurring when identifying anomalies like the model may miss anomalies if it predicts too well but also when it predicts poorly normal data may be misunderstood as anomalies. Having the balance between these two is essential for the accurate detection of anomalies in time series data. Detecting anomalies in time series data is more complicated than in a normal data, because of the time dependence and its non-stationary nature.

In this project, we propose a hybrid approach to anomaly detection in IoT time series data using a hybrid model integration method. Our approach includes using models such as Bi-LSTM + One-Class SVM, LSTM Autoencoder + Isolation Forest, and an ensemble model of DBSCAN, LOF, SVM to detect anomalies, visualize it, and compare the results through graphs. This method aims to enhance the security by making use of the unique strengths of each model.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024
DOI: 10.17148/IJARCCE.2024.13365

By using these diverse models, we aim to address the limitations of individual models and provide a more effective solution for anomaly detection in IoT in time series data. This method involving two models helps us to detect a wide range of anomalies which boosts the total discovery efficiency, to make IoT systems more secure and reliable. This project adds progress to the anomaly detecting abilities in IoT atmospheres preparing the means for even more efficient anomality detection services in the future. This work is motivated by the need to improve anomaly detection in IoT environments, where the undetected anomalies can cause severe problems. By the hybrid model method that we are proposing, we are aiming to enhance the accuracy and robustness of anomaly detection, leading to more reliable and secure IoT systems.

II. LITERATURE SURVEY

In paper [1] Anomaly Detection in Time Series Data of Sensex and Nifty50 With Keras, proposes an LSTM Autoencoder-based method for detecting anomalies in SENSEX and NIFTY50 stock market indices. This approach involves preprocessing the data, training the model, and using a threshold value to identify anomalies, aiding traders and investors in risk management.

In paper [2]Anomaly detection frameworks for outlier and pattern anomaly of time series in wireless sensor networks proposed hybrid approach to detect anomalies in wireless sensor networks. This hybrid approach leverages the strengths of CNN-LSTM for feature extraction from time series data and One-Class SVM for anomaly detection, offering a robust solution for anomaly detection in wireless sensor networks. This paper explored pattern anomaly detection based on raw time series data, highlighting the challenges and feature representation methods used for dimension reduction and noise filtration.

In paper[3] proposed T-DAD approach, a two-stage anomaly detection framework for manufacturing datasets, surpassing single-stage methods. T-DAD trains models in Stage I with operation cycle signals and in Stage II with sensor signals, showing superior performance. Its robustness is validated on other assembly process datasets. The T-DAD framework's performance may be influenced by the quality and quantity of the training data, as it requires a sufficient amount of labeled data to effectively train the models in Stage I and Stage II.

III. SCOPE AND METHODOLOGY

Aim of the project:The project entails a comprehensive exploration and implementation of a hybrid anomaly detection system tailored specifically for IoT time series data obtained from the Intel Berkeley Research Lab's 54 deployed sensors for unsupervised anomaly detection and UNSW -NB15 IoT weather board sensor data for supervised anomaly detection. This hybrid system combines various anomaly detection methodologies, such as One-Class SVM, Bi-LSTM, LSTM Autoencoder, Isolation Forest, DBSCAN, LOF, SVM, and a statistical approach.

In addition to model integration, the project will delve into the intricacies of preprocessing techniques optimized for IoT data streams, considering factors like missing data handling, feature selection, and normalization methods tailored to the dynamic nature of sensor data.The project's scope also extends to real-world applicability, emphasizing the deployment and validation of the hybrid anomaly detection system within the Intel Berkeley Research Lab environment. This deployment phase will involve rigorous testing under varying conditions to ascertain the system's robustness, scalability, and adaptability to dynamic IoT environments. Moreover, the project aims to develop intuitive visualization tools and interpretability techniques to facilitate stakeholders' understanding of anomaly detection results and actionable insights derived from the system.

1. Architecture Diagram

An architecture diagram is a visual representation of the structure and components of a system or model. It provides a high-level overview of how the system or model is organized and how its different components interact with each other. Architecture diagrams are commonly used in the field of computer science, machine learning, and deep learning to illustrate the design and structure of complex systems or models. Architectural design can also involve considering non-functional requirements such as performance, reliability, scalability, and security, and determining the best way to meet those requirements. The architectural design of the proposed system revolves around a hybrid approach of machine learning with deep learning model for anomaly detection. This architecture system involves using different Hybrid Anomaly Detection Models on the selected dataset which is a labelled or unlabelled dataset of IoT weather board sensor dataset. The data is chosen by the user then tested on the 3 models

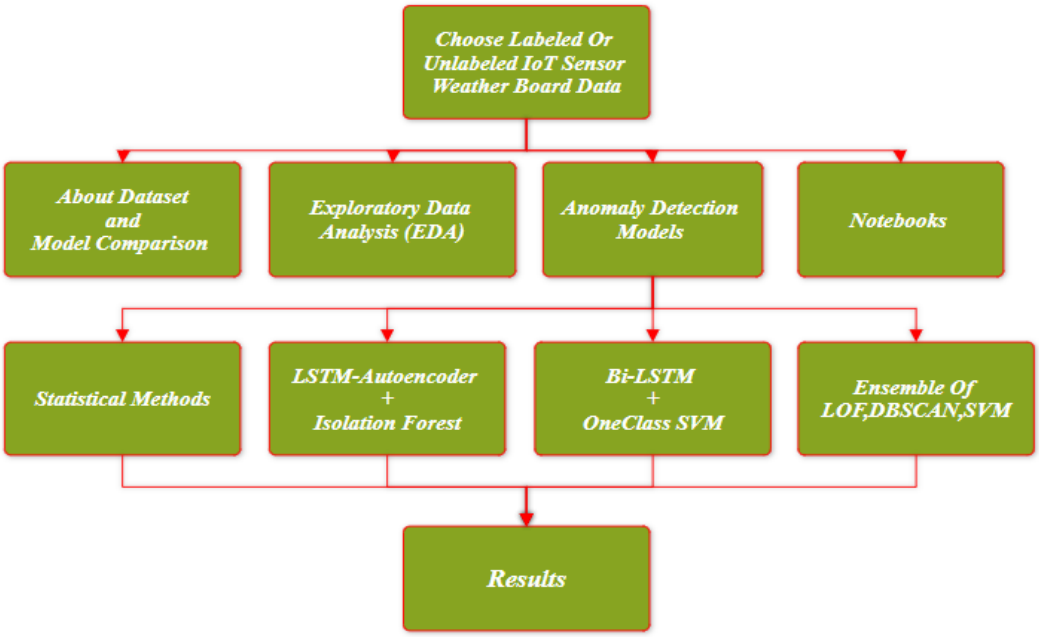
Bi-LSTM + One Class SVM , LSTM-Autoencoder + Isolation Forest , Ensemble of LOF, DBSCAN,SVM. Also Statistical approach is used with ARIMA, Z-score, IQR, modified Z-Score in finding anomalies .

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13365



2. Proposed System:

2.1 Bi LSTM + One Class SVM:

The proposed system integrates a Bi-directional Long Short-Term Memory (Bi-LSTM) autoencoder with a One-Class Support Vector Machine (SVM) for anomaly detection in time series data, specifically in IoT environments. The Bi-LSTM autoencoder learns temporal patterns in the data and reconstructs it, while the One-Class SVM detects anomalies based on the reconstruction error. This hybrid approach combines the strengths of deep learning and traditional machine learning to improve anomaly detection performance. We have used 2 layer Bi-LSTM architecture model with sigmoid activation function with output layer Dense. We used adam optimizer with loss function of binary_crossentropy. Training the Bi-LSTM model with 20 epochs , batch size of 32 has given a better result in the anomaly detection.the reconstruction errors from the Bi-LSTM architecture will be reshaped for One Class SVM model . These errors are then trained with One Class SVM with nu parameter of 0.01 with kernel rbf and then predictions are made using One Class SVM.

Out of 24260 Anomalies in the training set the model has been able to achieve 23610 anomalies which is a great result to achieve , accuracy of 96.34% is achieved in this model.

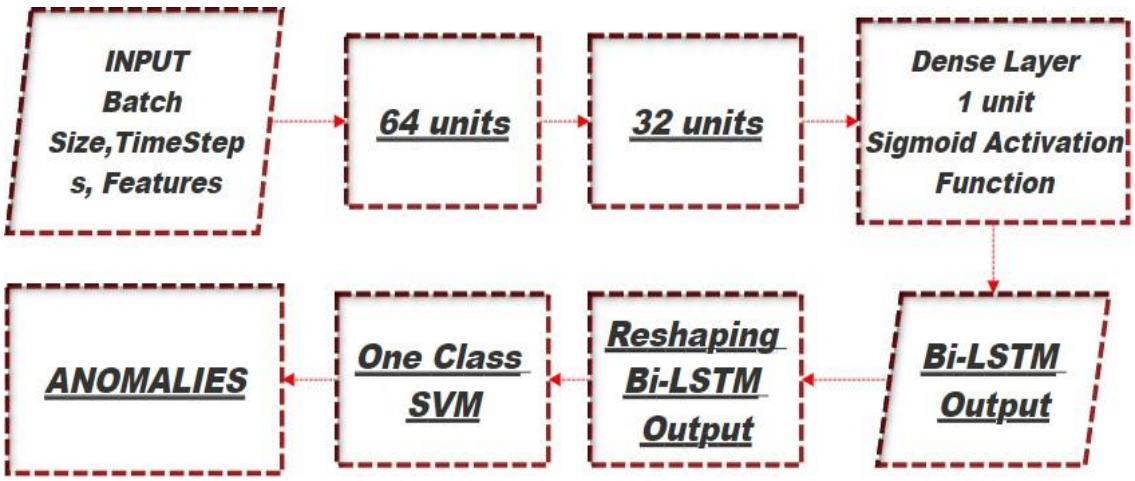


fig 1.1 Bi-LSTM + One Class SVM

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13365

2.2 LSTM Autoencoder + Isolation Forest:

The proposed system integrates an LSTM (Long Short-Term Memory) autoencoder with an Isolation Forest for anomaly detection in time series data, particularly in IoT environments. The LSTM autoencoder learns the temporal dependencies in the data and reconstructs it, while the Isolation Forest identifies anomalies based on the isolation of data points. This hybrid approach combines the strengths of LSTM in capturing complex temporal patterns and Isolation Forest in efficiently detecting anomalies, aiming to improve the accuracy and efficiency of anomaly detection in IoT environments. LSTM -Autoencoder architecture has layers of 4 also with a Repeat vector and Time Distributed with Dense layer for output. Relu activation function is used in this architecture with adam optimizer and loss function oof Mse. Training this model with epoch of 40 , batch size 64 the architecture is improved to sequence the time series input. The Isolation forest model is trained for detecting anomalies in the sequence from LSTM-Autoencoder . Since the highly imbalanced data the contamination factor of Isolation forest is kept at max of 0.5.This model is not able to achieve expected result in detecting anomalies in the whole dataset. Out of 24260 anomalies this model is only able to achieve 19560 anomalies, out of which the true positive count is about 15411 anomalies . The accuracy level of 90.8% is achieved

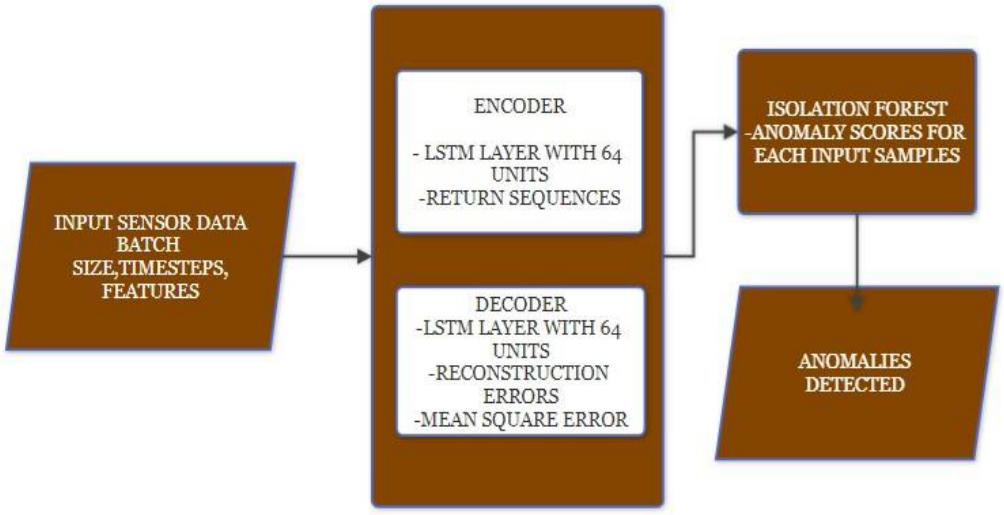


fig 1.2 LSTM Autoencoder + Isolation Forest

2.3 Ensemble Model (LOF, DBSCAN and SVM):

The proposed system utilizes an ensemble model consisting of Local Outlier Factor (LOF), DBSCAN (Density-Based Spatial Clustering of Applications with Noise), and Support Vector Machine (SVM) for anomaly detection in time series data, especially in IoT settings. The ensemble combines the anomaly detection results from each individual model to improve overall detection accuracy and robustness.

LOF identifies outliers based on local density deviation, DBSCAN clusters closely packed points and identifies outliers as points that are not part of any cluster, and SVM separates data points into normal and anomalous based on a hyperplane. By combining these diverse techniques, the ensemble model aims to provide more comprehensive and reliable anomaly detection in IoT environment. In DBSCAN with eps of 0.5 and min samples 5, Local Outlier Factor with n neighbors 20 and SVM with nu of 0.01.Ensemble scores from all the 3 models are extracted and then used up in the ensemble model. Threshold of 0.1 is considered as anomaly.

The mean of these anomaly scores are considered later up for the Ensemble Anomaly detection model. With this ensemble model we have achieved greater result with accuracy of 97.64 % with precision 1.0 indicating that anomalies detected by the model are correctly labelled as anomalies. The count of true positives is about 23688 which shows that this ensemble model is better in detecting anomalies. This is best model compared to the other two models in this project. However this is ensemble model so its computationally cost consumption.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024
DOI: 10.17148/IJARCCE.2024.13365

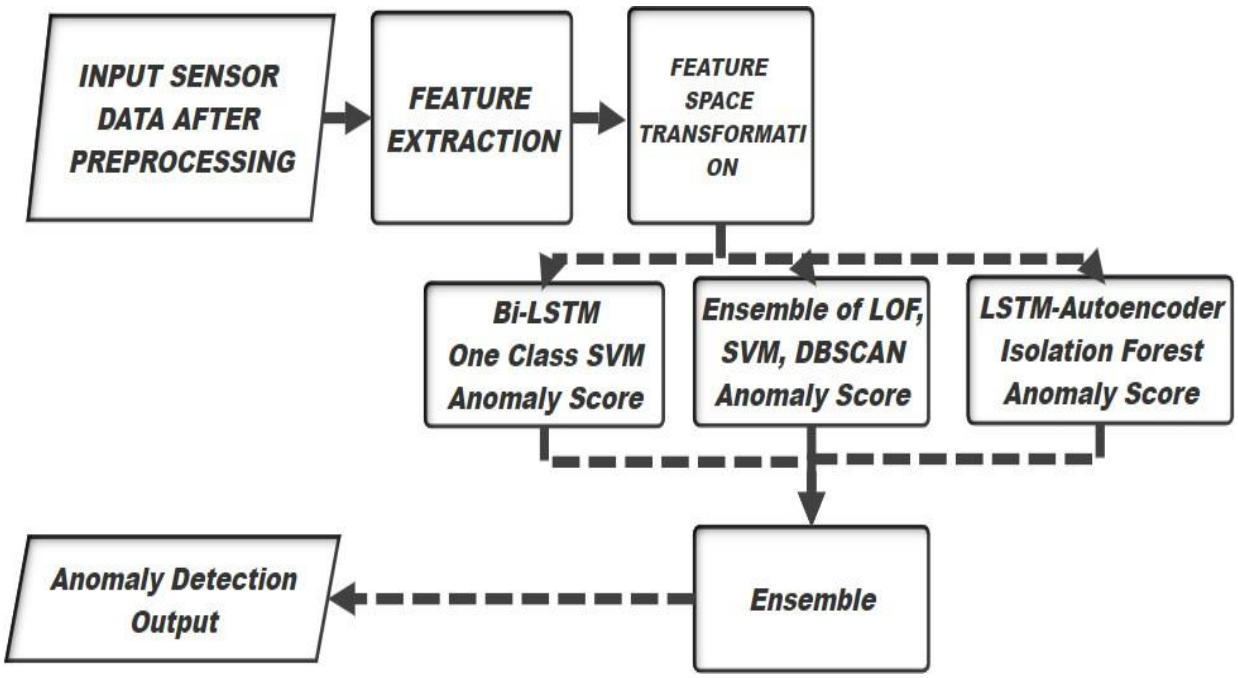


Fig 1.3 Ensemble Method

IV. CONCLUSION

In conclusion, this study explored various anomaly detection approaches, including statistical methods and deep learning models, for detecting anomalies in IoT weather data. The results showed that deep learning models, particularly the Bi-LSTM and LSTM autoencoder, outperformed traditional statistical methods such as z-score, IQR, MAD, and ARIMA, as well as other machine learning models like DBSCAN, LOF, and One-Class SVM. These deep learning models demonstrated high precision, recall, and F1-score, indicating their effectiveness in detecting anomalies in complex time-series data.

As conclusion it is found that out of the three Hybrid models we have used , the Bi-LSTM + One Class SVM and Ensemble of LOF , DBSCAN , SVM has been showing expected result .Our study has provided valuable insights into the application of machine learning and deep learning models for anomaly detection in complex datasets.

The Bi-LSTM + One-Class SVM model and the ensemble model have demonstrated superior performance, achieving accuracies of 96% and 97%, respectively, compared to the LSTM-Autoencoder + Isolation Forest model, which achieved 90% accuracy. While deep learning approaches can be powerful, they may not always outperform traditional machine learning algorithms. Overall, this study contributes to the growing body of literature on anomaly detection and provides a solid foundation for future research in this area.

REFERENCES

[1]. Dhruvil Shah,Soham Khade, Sudesh Pawar, "Anomaly Detection in Time Series Data of Sensex and Nifty50 With Keras ",In 2022, *International Conference on Emerging Smart Computing and Informatics (ESCI)* .
[2]. Gao, C., Chen, Y., Wang, Z., Xia, H. and Lv, N.,"Anomaly detection framework for outlier and pattern anomaly of time series in wireless sensor networks"In 2020 *International Conference on Networking and Network Applications (NaNA).*
[3]. Kyeong-Joong Jeong, Jin-Duk Park, Kyusoon Hwang, Seong-Lyun Kim, and Won-Yong Shin, "Two-Stage Deep Anomaly Detection With Heterogeneous Time Series Data ",In January 2022, *IEEE Access,VOLUME 10,*
[4]. Marco Pota, "Real-time anomaly detection on time series of industrial furnaces:A comparison of autoencoder architectures",In September 2023 *Engineering applications of Artificial Intelligence,Elsevier.*
[5]. Pengfie Liu,Elsevier, "Arrhythmia classification of LSTM autoencoder based on time series anomaly detection ",In January 2022 *Biomedical Signal Processing and Control Volume 71, Researchgate.*

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13365

[6]. Zhiwei Ji, Jiaheng Gong ,and Jiarui Feng, "A New Benchmark Using a Novel Deep Learning Approach",In July 2021, *Hindawi*.

[7]. Chunyong Yin, Sun Zhang, Jin Wang and Neal N. Xiong, "Anomaly Detection Based on Convolutional Recurrent Autoencoder for IoT Time Series",In May 2020,*IEEE Transactions on systems, man, and cybernetics: Systems*.

[8]. Fatma Taher , Mohamed Elhoseny, Mohammed K. Hassan , and Ibrahim M. El-Hasnony, "A Novel Tunicate Swarm Algorithm With Hybrid Deep Learning Enabled Attack Detection for Secure IoT Environment",In December 2022,*Volume 10,IEEE.*

[9]. Stephen Githinji, Ciira wa Maina, "Anomaly Detection on Time series Sensor Data Using Deep LSTM-Autoencoder",In 2023, *IEEE Africon.*

[10]. Lejla Begic Fazlic, Ahmed Halawa, "A Novel Hybrid Methodology for Anomaly Detection in Time Series, In June 2022, I*nternational Journal of Computational Intelligence Systems,Springer*.