iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: III Month of publication: March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78559

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

AI-Driven Anomaly Detection in Big Data Streams: A Case Study on IoT Healthcare ECG Data

Sreerasmi V M[1], Jisna C K[2], Sajana T[3], Anagha P[4]
Assistant Professor, Computer Science Department, Chinmaya Institute of Technology, Kannur, Kerala, India

Abstract: Early detection of cardiac abnormalities is crucial for effective healthcare intervention, especially with the increasing adoption of IoT-based wearable devices. This study presents an AI-driven approach for anomaly detection in ECG time-series data using unsupervised learning techniques, namely Isolation Forest and Local Outlier Factor (LOF). The analysis is conducted on publicly available ECG datasets, including a standard ECG time-series dataset and the PTB Diagnostic ECG Database (PTBDB). The models are evaluated based on anomaly detection capability and comparative performance. Experimental results show that Isolation Forest detects a slightly higher number of anomalies, while LOF effectively captures local density variations. The study demonstrates that lightweight unsupervised models can be integrated into IoT healthcare systems for real-time cardiac monitoring. Future work includes extending the approach using deep learning models and real-time streaming data analytics.
Keywords: ECG signal, anomaly detection, Isolation Forest, Local Outlier Factor, IoT healthcare, big data analytics, PTBDB, time-series analysis.

I. INTRODUCTION

The continuous growth of IoT-based medical devices has led to an explosion of real-time health monitoring data, particularly ECG signals, which are vital for early cardiac abnormality detection [2]. Anomalies in ECG patterns can indicate life-threatening conditions such as arrhythmias or cardiac arrest [2]. Traditional monitoring systems rely heavily on clinical interpretation and labeled data, which limits scalability in remote or continuous care setups. Leveraging AI-based unsupervised anomaly detection techniques allows healthcare systems to flag potential risks automatically without the need for extensive labeled datasets. [2]

This study explores two unsupervised learning algorithms, Isolation Forest [6] and Local Outlier Factor (LOF) [1], applied to ECG data to detect anomalies. The analysis focuses on publicly available datasets, including time-series ECG signals obtained from Kaggle [5] and the PTBDB dataset hosted on PhysioNet [3].

A. Contributions of the Study

This work makes the following contributions:

1) Applies two unsupervised anomaly detection algorithms (Isolation Forest and LOF) on real-world ECG datasets.
2) Performs a comparative analysis of both models based on anomaly detection results.
3) Provides visualization of detected anomalies in ECG signals.
4) Demonstrates the feasibility of deploying lightweight AI models in IoT-based healthcare systems.

II. RELATED WORK

Numerous studies have applied machine learning and AI techniques to ECG anomaly detection [2]. Traditional methods include supervised learning models like SVM (Support Vector machines) and Decision Trees, requiring labeled data. Recently, researchers have adopted unsupervised algorithms such as One-Class SVM, DBSCAN, and Autoencoders for ECG analysis [2].

Breunig et al. introduced LOF for density-based outlier detection [1], which has since been adapted for time-series health data. Liu et al. proposed Isolation Forest, an ensemble method known for its efficiency with high- dimensional data [6]. Both algorithms have proven effective in anomaly detection tasks in various domains, including finance, cybersecurity, and healthcare [2]

A study titled “Anomaly detection based on Artificial Intelligence of Things: A Systematic Literature Mapping” provided a broad review of anomaly detection strategies in AIoT healthcare applications but emphasized the need for practical case studies applying unsupervised algorithms on real ECG datasets [2].

Many of the prior works lack comprehensive visualization and practical anomaly quantification, limiting their applicability in real-time IoT healthcare systems. However, comparative studies focusing on these models applied to ECG time-series datasets, especially using open-source IoT healthcare datasets, remain limited.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

This work addresses that gap. This study addresses this gap by applying both Isolation Forest and LOF algorithms on ECG time-series and PTBDB datasets, presenting a comparative analysis, visualization of anomalies, and insights into the practical deployment of such models in IoT-enabled cardiac monitoring.

A. Related *Work Summary*

1) Zhang et al. (2020) applied Support Vector Machine (SVM) and Decision Tree classifiers to the MIT-BIH Arrhythmia dataset, achieving good accuracy in ECG classification. However, *their approach relied entirely on supervised methods,* necessitating large volumes of labeled data for effective performance.
2) Kumar et al. (2021) utilized the Isolation Forest algorithm for anomaly detection in ECG signals collected from IoT-based wearable devices. While their study successfully identified outliers, it lacked a comparative analysis with other unsupervised learning algorithms, limiting its generalizability.
3) Khan et al. (2022) implemented the Local Outlier Factor (LOF) technique on the PTB Diagnostic ECG Database (PTBDB) and effectively detected local anomalies within the dataset. Despite this, their study did not include a comparison with Isolation Forest and *did not test on IoT-based streaming data.*

A systematic literature mapping by an anonymous group (2023) reviewed anomaly detection approaches in AIoT environments using multiple IoT datasets. While comprehensive in scope, the study offered *no practical case study implementation and lacked a specific focus on ECG-based anomaly detection.*

The proposed study (2026) applies both Isolation Forest and Local Outlier Factor algorithms on ECG time-series data and the PTBDB dataset. This work uniquely detects and visualizes anomaly points in ECG signals, performing a comparative evaluation of both models. It stands out by combining both algorithms on real-world ECG datasets with practical anomaly detection implementation, addressing gaps in prior research.

III. METHODOLOGY

A. Dataset Description

1) ECG Time-Series Dataset: Real-time ECG signal data [5] with fields ‘time’ and ‘ecg_value’.
2) PTB Diagnostic ECG Database (PTBDB): Publicly available clinical ECG data containing multiple patient records. It is hosted on PhysioNet [3], originally created by the Physikalisch-Technische Bundesanstalt (PTB), the National Metrology Institute of Germany.

B. Preprocessing

1) Converted data into CSV format.
2) Normalized ECG values.
3) Removed duplicates and handled missing values.

C. Algorithms Applied

1) Isolation Forest (IF)

Detects anomalies by isolating data points using random splits [6]. The anomaly score is computed based on the path length:

h(x)=average path length of point x.

Shorter path lengths indicate anomalies.

2) Local Outlier Factor (LOF)

Identifies local density deviations to detect outliers [1].

LOF(x)= Average Local Reachability Density of Neighbors/Local Reachability Density of x

Values greater than 1 indicate anomalies.

Both models were implemented using the scikit-learn Python library [7].

D. Implementation Environment

1) Python 3.x
2) Pycharm IDE
3) Libraries: pandas, numpy, matplotlib, scikit- learn

IV. RESULT AND DISCUSSION

Isolation Forest detected a slightly higher number of anomalies compared to LOF in both datasets. Visualization of anomalies showed noticeable spikes or dips in ECG signals at anomalous points.

A. Isolation Forest Output

Anomalies detected: 203 (ECG time-series), 179 (PTBDB)

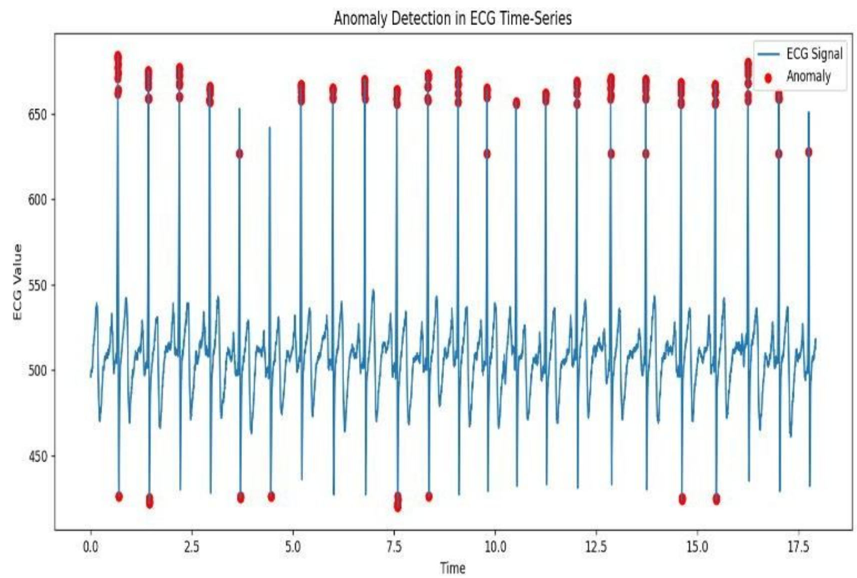
B. LOF Output

Anomalies detected: 179 (ECG time-series)

Both algorithms effectively highlighted potential risk points in ECG data, confirming their reliability for unsupervised anomaly detection in healthcare.

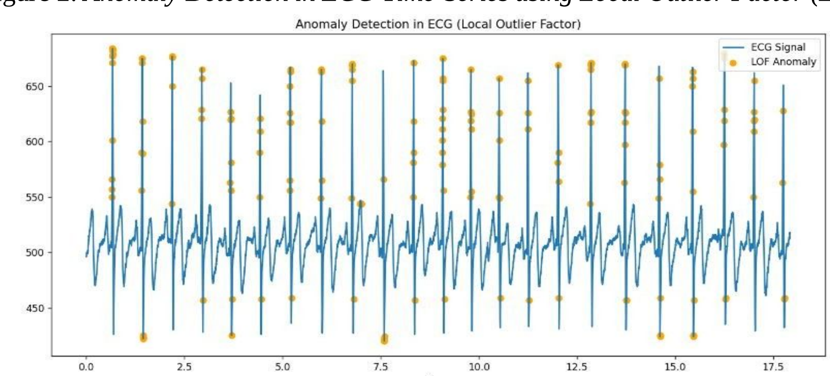
C. Result Visualization

Figure 1: Anomaly Detection in ECG Time-Series using Isolation Forest.



Explanation: This plot displays ECG time-series data (blue line) alongside anomalies (red dots) detected by Isolation Forest. The algorithm effectively isolated sharp spikes and deviations, which may indicate abnormal cardiac activity

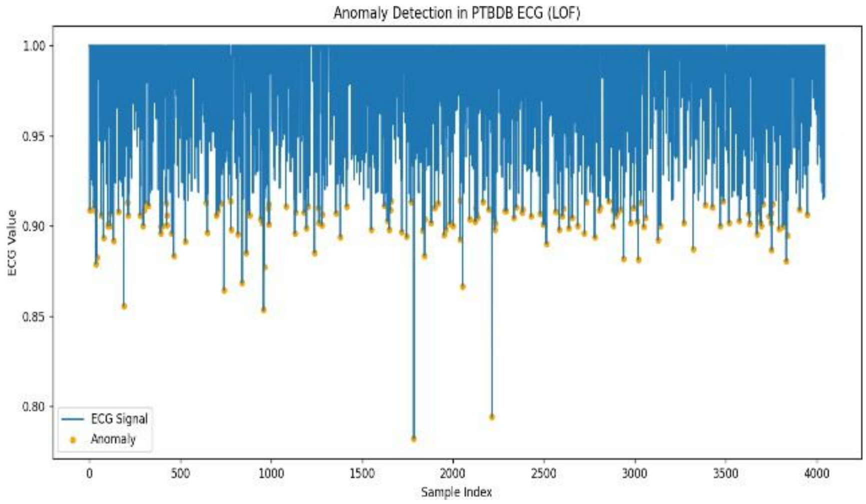
Figure 2: *Anomaly Detection in ECG Time-Series using Local Outlier Factor (LOF).*



Explanation: This figure shows the ECG signal in blue, with anomalies detected by LOF marked as orange dots. LOF detects local density-based deviations, highlighting points where ECG values significantly differ from their neighbors, typically aligning with spikes and dips.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

Figure3: Anomaly Detection in PTBDB ECG Data using Local Outlier Factor (LOF).



Explanation: This visualization presents PTBDB ECG data analyzed with LOF. The orange dots represent detected anomalies, generally occurring at sharp downward spikes or irregular values in the normalized ECG signal. It confirms LOF's suitability for clinical ECG anomaly detection.

D. *Comparative Results*

Algorithm	Dataset	Anomalies Detected	Observation
Isolation Forest	ECG Time-Series	203	Detects more global anomalies
Isolation Forest	PTBDB	179	Consistent performance
LOF	ECG Time-Series	179	Detects local density anomalies

V. CONCLUSION AND FUTURE WORKS

A. *Conclusion*

This study presented a comparative analysis of two unsupervised anomaly detection algorithms, Isolation Forest and Local Outlier Factor, applied to ECG datasets. Both models successfully identified anomalous patterns in ECG signals without requiring labelled data. Isolation Forest demonstrated slightly higher sensitivity, while LOF effectively captured local variations. The findings highlight the potential of integrating lightweight AI models into IoT-based healthcare systems for continuous and real-time cardiac monitoring. The analysis revealed that:

1) Isolation Forest consistently detected a slightly higher number of anomalies compared to LOF.
2) Both algorithms successfully identified significant outlier points in ECG signals, supporting their reliability in healthcare anomaly detection scenarios.
3) This practical case study bridges a gap in existing literature by comparatively evaluating multiple unsupervised techniques on real healthcare datasets.

The findings suggest that integrating these models into IoT-based wearable healthcare devices could enable early detection of critical heart conditions, improving patient safety and proactive clinical intervention proactive clinical intervention.

B. *Future Work*

1) Explore deep learning models like LSTM Autoencoders. An LSTM (Long Short-Term Memory) Autoencoder [4] is a type of deep learning model designed to compress (encode) sequential or time-series data into a smaller representation and then reconstruct (decode) it back to the original.
2) Implement real-time detection in IoT streaming environments.
3) Integrate multi-channel ECG and patient metadata for improved accuracy

REFERENCES

[1] Breunig, M.M., Kriegel, H.P., Ng, R.T., and Sander, J. LOF: Identifying Density-Based Local Outliers. Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 29(2), 93–104, 2000. https://doi.org/10.1145/335191.335388

[2] Chandola, V., Banerjee, A., and Kumar, V. Anomaly Detection: A Survey. ACM Computing Surveys (CSUR), 41(3), 1–58, 2009. https://doi.org/10.1145/1541880.1541882

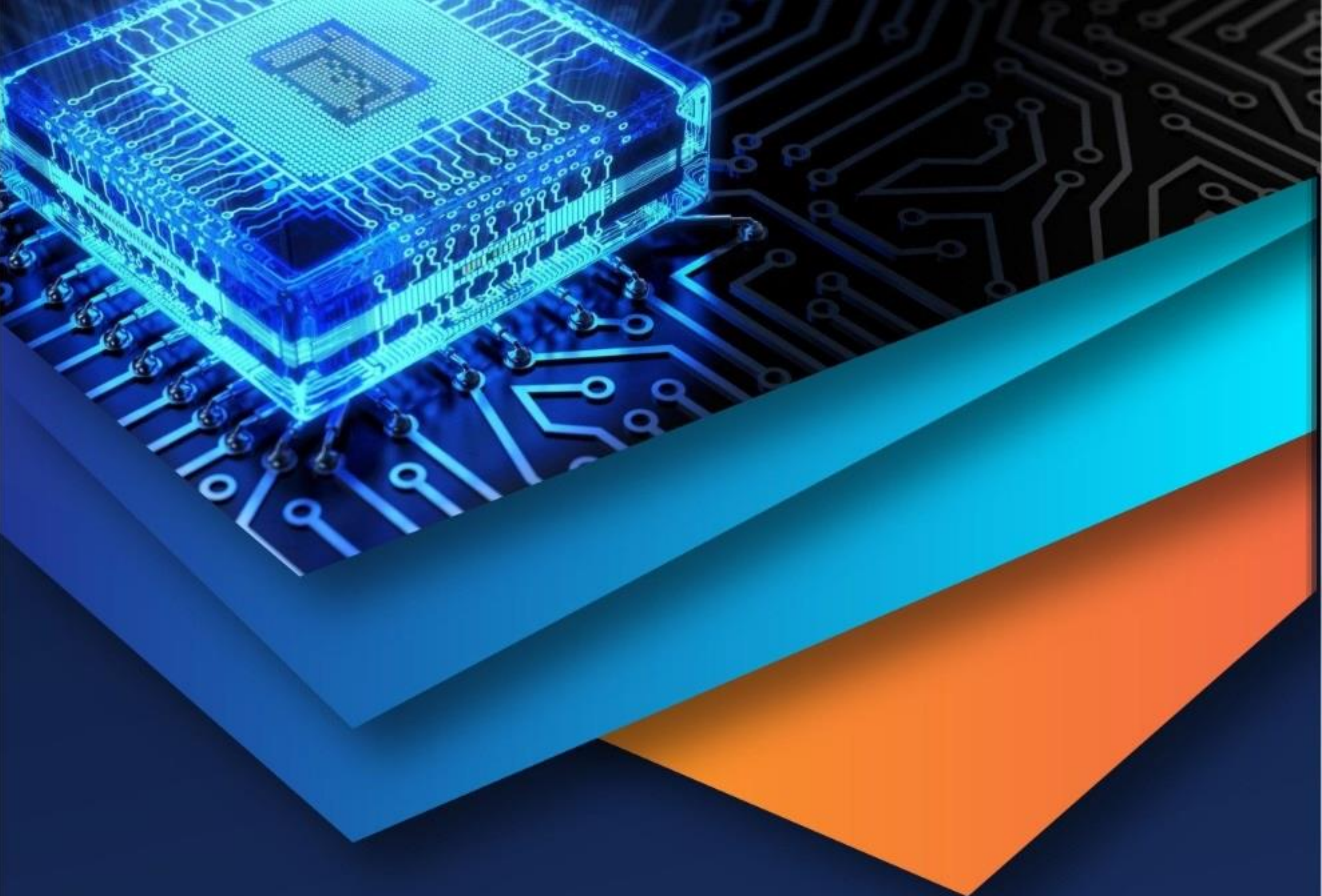
[3] Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B.,Peng, C.K., and Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation, 101(23), e215–e220, 2000. https://doi.org/10.1161/01.CIR.101.23.e215

[4] Kachuee, M., Fazeli, S., and Sarrafzadeh, M. ECG Heartbeat Classification: A Deep Transferable Representation. Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI), 443–444, 2018. https://doi.org/10.1109/ICHI.2018.00091

[5] Kaggle. ECG Time Series Data for Anomaly Detection [Data set]. 2024. https://www.kaggle.com/dataset-url

[6] Liu, F.T., Ting, K.M., and Zhou, Z.H. Isolation Forest. Proceedings of the 2008 IEEE International Conference on Data Mining, 413–422, 2008. https://doi.org/10.1109/ICDM.2008.17

[7] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830, 2011. https://jmlr.org/papers/v12/pedregosa11a.html



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY