

# Anomaly Detection in IoT Based Healthcare Monitoring Systems Using KNN and DBSCAN Machine Learning Techniques

Joslin Gadde

Student ID: 20076078

Nada Said

Student ID :20024717

Namuna Upreti

Student ID: 20068236

Tanvi Prajapati

Student ID: 20018024

Damilola Abosede

Student ID: 20053801

**Abstract**—The growth of Internet of Things (IoT) technologies in healthcare has enabled continuous monitoring of physiological parameters through wearable and embedded sensors. As healthcare systems increasingly adopt data driven monitoring, anomaly detection has become essential for identifying deviations that may indicate health deterioration, sensor malfunctions, or device failures, thereby improving reliability and patient safety. The aim of this study is to detect anomalies in IoT healthcare sensor data using K Nearest Neighbours (KNN) and Density Based Spatial Clustering of Applications with Noise (DBSCAN), and to evaluate their comparative performance in identifying abnormal physiological patterns. KNN based anomaly scoring offered an intuitive way to capture point level deviations in physiological measurements, while the Local Outlier Factor (LOF) enhanced sensitivity to contextual and local density changes. DBSCAN detected density anomalies and sensor noise in unstructured sensor streams, providing complementary anomaly detection signals, point wise, density aware, and cluster level, improving reliability compared with any single model. The results suggest that such hybrid unsupervised methods have strong potential for supporting trustworthy IoT based clinical monitoring pipelines..

**Index Terms**—Wearable health sensors, Anomaly detection, K nearest neighbours (KNN), DBSCAN clustering, Sensor data reliability

## I. INTRODUCTION

### A. Background

The growth of Internet of Things (IoT) technologies in healthcare has enabled continuous monitoring of physiological parameters through wearable and embedded sensors. These devices capture and transmit real time data such as heart rate, temperature, and oxygen saturation, enabling early detection of abnormal physiological patterns and supporting proactive medical intervention [1]. As healthcare systems increasingly adopt data driven monitoring, anomaly detection has become essential for identifying deviations that may indicate health deterioration, sensor malfunctions, or device failures, thereby improving reliability and patient safety.

### B. Related Works

Recent work highlights the growing use of machine learning for anomaly detection in IoT and IoMT systems. Khan et al. proposed a deep learning-based framework for identifying irregularities in medical sensor data, improving intelligent healthcare monitoring [1]. Hannah et al. evaluated algorithms such as KNN, SVM, and clustering methods for anomaly

detection in IoT networks, demonstrating techniques that are transferable to physiological data analysis [2].

Classical methods remain widely used in IoT healthcare: KNN captures distance-based outliers, while DBSCAN detects density anomalies and sensor noise particularly valuable when labelled data are limited [4]. The CRISP-DM framework continues to guide structured data preparation, modelling, and evaluation in such applications.

### C. Research Gap

Despite advances, key gaps persist. Comparative evaluations of KNN vs. DBSCAN for anomaly detection in IoT healthcare data are limited, as most studies prioritise deep learning or single model analyses. Interpretability is also under addressed; visual explanation methods such as PCA plots or anomaly score distributions are rarely applied, reducing clinical usability [8]. Additionally, real world IoT challenges sensor noise, missing values, irregular sampling, and device constraints are often overlooked, creating a disconnect between algorithmic research and practical healthcare deployment [6].

### D. Aim and Objectives

The aim of this study is to detect anomalies in IoT healthcare sensor data using KNN and DBSCAN, and to evaluate their comparative performance in identifying abnormal physiological patterns. Objectives:

- Import, clean, and preprocess IoT healthcare sensor data.
- Conduct statistical analysis and visualisation to understand overall data behaviour.
- Implement KNN anomaly detection using the PyOD library.
- Implement DBSCAN clustering based anomaly detection.
- Compare both methods using PCA visualisation and anomaly scoring.
- Evaluate and interpret anomaly results.
- Discuss ethical, legal, and social implications of ML in healthcare.
- Apply the CRISP DM lifecycle throughout the project.

## II. METHODS

### A. Dataset Description and Acquisition

The dataset simulated physiological and device data from wearable IoT sensors, sourced from Kaggle [10]. Each record

includes Patient ID, Timestamp, and Sensor ID, capturing vital signs, device metrics, and clinical reference values [10]. It was imported into Python for cleaning, transformation, and anomaly-detection preparation, with all numerical features subsequently standardised [9]. To support anomaly detection, all numerical features were later transformed using the standardization formula:

$$x' = x \frac{\mu}{\sigma}$$

ensuring comparability and consistent scaling for all physiological measurements.

### B. Data Understanding

The dataset used in this study is an open source, simulated IoT healthcare dataset containing continuous measurements of key patient vitals. The main columns relevant to this work include:

- Temperature ( $^{\circ}\text{C}$ ) – body temperature
- BP (Blood Pressure) – systolic or combined BP indicator
- HR (Heart Rate) – beats per minute
- Battery – IoT device battery level (%)

As is common with real world IoT data streams, the dataset exhibited several typical issues:

- Missing values caused by transmission loss or temporary sensor failure
- Noise in sensor readings due to environmental interference
- Sensor drift, where values slowly deviate from true measurements over time [20]

Understanding these data characteristics is essential for developing robust anomaly detection models, since anomalies may result from either real clinical events or faulty sensor behavior[8]. To quantify noise and drift, distance based anomaly measures such as Euclidean distance were later applied:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}$$

This mathematical formulation supports both KNN and DBSCAN methods used in later stages.

### C. Data Preparation

Data preparation involved cleaning, feature selection, and standardisation to make the dataset suitable for machine learning. Only relevant numeric sensor fields such as temperature, humidity, pressure, voltage, and current were retained to reduce noise and dimensionality [11]. Converting these values to consistent numerical formats streamlined modelling and supported effective anomaly detection [1]. Once the relevant sensor columns were identified, all values within those columns were converted to numerical types using the following transformation:

This step served two main purposes:

- 1) Standardization of Data Types: Sensor datasets often contain values stored as strings due to formatting issues, delimiters, or data logging inconsistencies. Converting all entries to numeric ensured that the columns could be

processed correctly by statistical and machine learning methods.

### 2) Handling Invalid or Corrupted Values:

Using errors='coerce' ensured that any non numeric entries such as text, missing symbols, or corrupted readings were converted to NaN, preventing runtime errors and enabling systematic handling of missing or inconsistent values. This step established a clean and standardised numerical dataset, forming a reliable foundation for subsequent machine learning analysis

### D. Handling Missing Values

Missing values were addressed using mean imputation, a common and effective approach for continuous physiological and sensor based data. This method replaces each missing entry with the mean value of its corresponding feature, ensuring consistency while minimizing distortion of the underlying distribution.

The following operation was used:

$$x_{ij}^* = \begin{cases} x_{ij}, & \text{if } x_{ij} \text{ is not missing} \\ \frac{1}{n} \sum_{k=1}^n x_{kj}, & \text{if } x_{ij} \text{ is missing} \end{cases} \quad \text{data} =$$

data.fillna(data.mean()) This approach offers several advantages:

- Preserves Dataset Size: No rows are removed, preventing data loss and maintaining the integrity of the time series or sampling frequency.
- Simple and Robust: Mean substitution performs well when missingness is relatively low and the data are symmetrically distributed.
- Algorithm Compatibility: KNN and DBSCAN cannot handle missing values natively; therefore, imputation ensures the dataset is complete and ready for distance based computations [21].

By filling in missing entries, the dataset becomes fully usable for subsequent anomaly detection tasks without introducing structural gaps.

### E. Scaling

KNN and DBSCAN rely on distance calculations, so features with larger numeric ranges (e.g., battery level) can dominate smaller ones (e.g., heart-rate variability), skewing anomaly detection [22]. To prevent this, features were standardised using StandardScaler to have zero mean and unit variance:

$$x' = \frac{x - \mu}{\sigma}$$

scaler = StandardScaler()

data\_scaled = scaler.fit\_transform(data)

This transformation provides:

- Equal Weighting Across Features: No single variable disproportionately influences distance metrics.
- Improved Model Performance: Distance based algorithms operate more effectively when features are standardized.

- Enhanced Anomaly Detection Accuracy: Subtle deviations in physiological signals become more detectable once scale driven distortions are removed.

Overall, standardization ensures that all sensor measurements contribute uniformly to the clustering and classification steps involved in detecting anomalies.

#### F. Modelling

This stage applied two classical anomaly-detection algorithms K-Nearest Neighbours (KNN) and DBSCAN followed by Principal Component Analysis (PCA) for visual interpretation. The methods are complementary: KNN flags distance-based deviations, while DBSCAN identifies points that fall outside dense clusters. Both are well suited to IoT healthcare data, where abnormal physiological readings or sensor faults often appear as distance- or density-based outliers [23].

*1) KNN Based Anomaly Detection (PyOD Implementation):* The KNN algorithm identifies anomalies by measuring how far a data point lies from its k nearest neighbours. Points with unusually large neighbour distances are considered anomalous [9].

##### Distance Computation

KNN uses Euclidean distance to compute similarity between each pair of observations:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}$$

*2) DBSCAN Density Based Anomaly Detection:* DBSCAN is a density based clustering algorithm that identifies areas of high point concentration and labels points in sparse regions as anomalies. This makes it well suited for IoT healthcare data, where anomalies often appear as isolated readings outside the normal physiological range or as the result of sensor malfunction [12].

DBSCAN requires two key parameters:

- eps (the radius defining a neighbourhood around each point)
- min samples (the minimum number of points needed to form a dense region)

In this study, eps = 1.2 and min samples = 10 were chosen after exploratory tests on the scaled dataset. These values allowed the algorithm to form meaningful clusters of normal behaviour while still identifying scattered points as anomalies.

DBSCAN labels outliers using the value 1, which was converted into a binary format for easier interpretation.

##### Implementation:

- dbscan = DBSCAN(eps=1.2, min samples=10)
- db labels = dbscan.fit predict(data scaled)
- df["DBSCAN Anomaly"] = (db labels == 1).astype(int)

DBSCAN is advantageous because it does not assume any specific data distribution. It is capable of detecting anomalies that form small clusters or appear as isolated points, making it complementary to the distance based approach used by KNN [12].

#### G. Evaluation

Since this is unsupervised, evaluation includes:

- PCA visual separation: Principal Component Analysis (PCA) reduced high-dimensional sensor data to two components, enabling visualisation of clusters representing normal patterns and highlighting significant deviations. PCA plots offered an interpretable view of model behaviour and supported qualitative comparison of KNN and DBSCAN anomaly detection [24].
- Frequency of anomalies: Anomaly counts were compared to assess algorithm sensitivity. KNN, a distance-based method, identifies fewer extreme deviations, while DBSCAN, using density-based clustering, detects more anomalies, especially in sparse regions [15]. Comparing frequencies helps reveal potential false negatives and false positives, critical in clinical settings [23]..
- Comparison of KNN vs DBSCAN: A comparative evaluation of KNN and DBSCAN was conducted based on several criteria:
  - Detection Sensitivity: Assessing the ability to identify extreme or subtle deviations from normal physiological patterns [16].
  - Robustness to Noise: Evaluating the algorithms' capacity to differentiate between sensor noise and genuine physiological anomalies.
  - Interpretability: Determining the ease with which detected anomalies can be visualised and interpreted in the clinical context.

This analysis elucidates the relative strengths and limitations of each methodology, informing the selection of appropriate models for deployment in real time healthcare monitoring systems.

- Correlation heatmap: A correlation heatmap was constructed to examine interdependencies among sensor variables. Strong correlations, such as those observed between systolic and diastolic blood pressure, provide a benchmark for physiological consistency. Anomalous observations that contradict expected correlations were flagged for further investigation, thereby reinforcing the validity of detected anomalies through domain informed contextual validation [24].

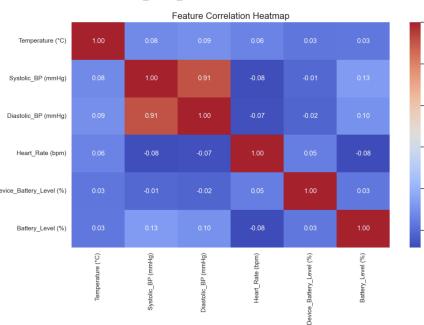


Fig. 1: Correlation HeatMap

- Domain interpretation (medical context): Anomalies were interpreted clinically and operationally. Sudden changes

in vital signs were viewed as potential indicators of acute health issues, while device-related anomalies such as rapid battery drops or irregular sensor readings suggested hardware or connectivity faults [11]. This dual interpretation ensures the results are both statistically valid and clinically meaningful for IoT healthcare monitoring.

### III. RESULTS

#### A. Statistical Overview

The initial step we then examined was the fundamental statistics of the vital signs, such as heart rate, blood pressure, temperature, and activity levels, to get an idea of what normal was in the data [2].

- The heart rate was largely in the normal range with some high spikes.
- Blood pressure had a broader distribution, and this indicates that there was more variance with some apparent peaks.
- Temperature was very stable with minimal anomalous values.
- Correlation heatmap was used to demonstrate the relationship among the variables. As anticipated, systolic and diastolic blood pressure were closely related.
- The rate of heart and activity was sometimes opposite. Temperature did not relate at all with some other variable which in fact makes it unique when something weird occurs.

#### B. Visualisations

**PCA Cluster Plots:** We employed PCA to minimize all the features to two components to make the dataset easier to visualize. The plot depicted a single cluster (normal data) with a few scattered points around the cluster (possibility of anomalies).

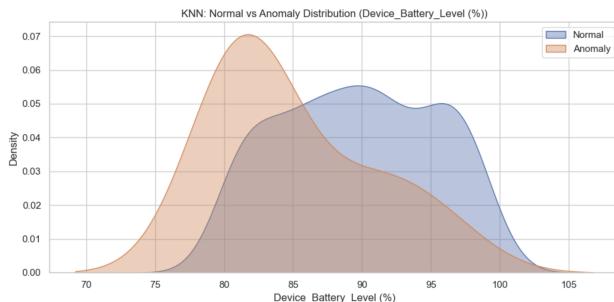


Fig. 2: KNN Normal vs Anomaly Distribution Battery Level

**KNN Anomaly Detection:** The KNN plot had normal points remaining clustered as outliers were dispersed primarily on the edges. KNN identified the apparent outliers which were the points that were obviously not part of the rest of the data. It was discriminating and it selected only the strangest.

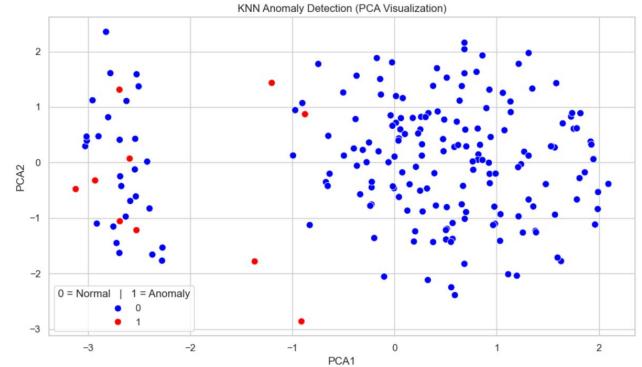


Fig. 3: KNN Anomaly Findings

**DBSCAN Anomaly Detection:** DBSCAN was considerably different. It counted much more points as abnormal cases, particularly at those points where there was a weak concentration of the data or sparseness.

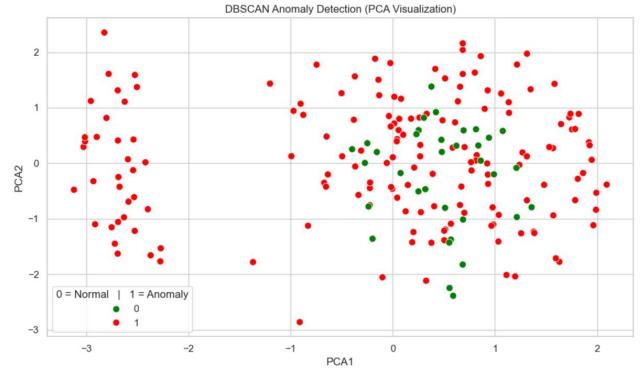


Fig. 4: DBSCAN Anomaly

This implies that DBSCAN will be able to spot smaller patterns or hidden patterns more effectively, but it will also falsely consider normal points as suspicious. Anomaly Count Comparison When comparing both methods: KNN scored only a few points as anomalies.

- DBSCAN marked many more.

This indicates that KNN is more stringent whereas DBSCAN is more sensitive to the density change.

#### C. Anomaly Findings

In the two algorithms, the following forms of anomalies were recognized:

- High Blood Pressure Spikes Acute rises in blood pressure, which were not in conformity to activity levels. These may be signs of stress or hypertension occurrences.
- Low Heart Rate Drops The occasional slowing of heart rate which was below normal. Despite being brief, they were evident breaks of the normal rhythm of the person.
- Critical Hardware Problems Sensor Battery or Hardware Problems. Some of the readings had sudden falls or irregular indications probably due to low battery or sensor transgression. DBSCAN was able to identify these with less difficulty since they interfere with local density patterns.

- Abnormal Tempering Fluctuations. Some of the temperature readings were out of the norm either high or low. These were noticeable in the data.

1) *Comparison of Methods:* KNN is excellent with self evident, solitary outliers and with high dimensions. Nevertheless, it tends to overlook minor problems within groups. DBSCAN is more effective in identifying significant anomalies in data that is dirty, or in data that is spatially skewed and clumped. It can also exaggerate, and more points can be scored as anomalies.

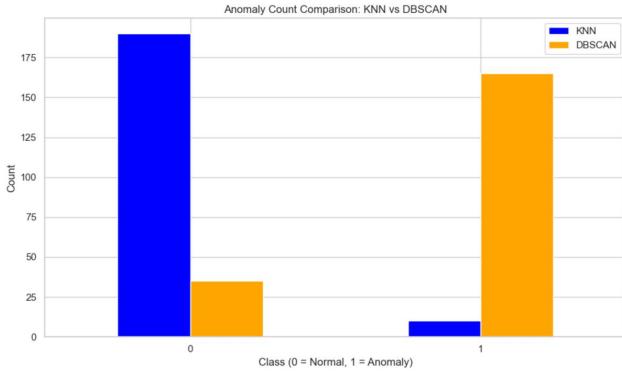


Fig. 5: KNN vs DBSCAN Anomaly findings

TABLE I: Comparison of Methods

| Metric                        | KNN                      | DBSCAN |
|-------------------------------|--------------------------|--------|
| Sensitivity                   | High for strong outliers | Medium |
| Handles Noise                 | Medium                   | High   |
| Works Well in High Dimensions | Yes                      | No     |
| Good for Clustered Data       | No                       | Yes    |

## IV. DISCUSSION

### A. Interpretation

Physiological anomalies can indicate early deterioration, including hypoxemia or arrhythmia. Density-based methods (KNN/LOF) detect deviations from a patient's baseline, while DBSCAN highlights points outside normal clusters. Studies show that automated vital-sign analysis can separate artefacts from true events when validated with clinician-labelled data [12].

Device-related anomalies are equally important. Battery depletion, charger faults, or connector damage can cause sudden dropouts or flat-lined readings that differ from physiological issues but still undermine monitoring reliability. Real-world recalls and incident reports highlight these failures as patient-safety risks, reinforcing the need to include device-health features in anomaly-detection pipelines [3], [4].

### B. Strengths Weaknesses

#### KNN:

- Works well for numeric features and provides an intuitive, instance based measure of neighbourhood similarity. [5]

- Weakness: sensitive to feature scaling and irrelevant/noisy features; therefore, standardisation or careful feature selection is needed for robust distance calculations. [6]

#### DBSCAN:

- Identifies natural clusters of normal behaviour and labels points in low density regions as anomalies without requiring a fixed number of clusters. [11]
- Weakness: requires careful tuning of eps and min samples (or equivalent) and can struggle in varying density contexts; preprocessing (scaling, dimensionality reduction) frequently improves stability. [7], [8]

### C. CRISP DM Reflection

Following CRISP-DM emphasised that each project phase contributed materially to final model accuracy. Business and data understanding clarified which anomalies were clinically relevant versus device artefact; data preparation (cleaning, imputation, scaling, feature engineering) directly improved KNN/LOF distance calculations and DBSCAN clustering; and iterative modelling and evaluation allowed tuning of hyperparameters with clinically meaningful metrics rather than only mathematical scores. Systematic reviews of CRISP-DM in practice reinforce that structured, iterative workflows reduce overfitting and support reproducible parameter selection in applied settings. [13]

### D. Ethical, Legal, Social, Professional Issues

Health data are special-category personal data requiring GDPR-compliant safeguards, including data minimisation and documented consent. UK guidance from the ICO and NHS Digital sets standards for secure handling [14], [15]. Clinical settings also demand confidentiality and traceable governance to maintain trust, as failure to meet these can hinder adoption even with strong model performance [11].

Bias is a major concern: if training data under-represent certain demographic groups or clinical conditions, detectors may perform poorly for those populations. Studies in health AI show that uneven data coverage leads to systematic errors, making stratified evaluation, re-sampling, and fairness-aware modelling necessary to reduce disparate impacts [17], [18].

Sustainability is also relevant. Large-scale IoT monitoring increases energy use and environmental impact. Techniques such as energy-efficient sensing, edge pre-processing, and duty-cycling can reduce transmission and battery consumption while preserving detection performance, supporting sustainable healthcare deployments [19].

Achieving algorithmic fairness in medicine requires both evaluation across demographic and clinical groups and mitigation through retraining, constraints, or post-processing. Reviews emphasise that fairness relies on better data, transparency, and domain-informed design rather than generic fairness tools alone [18], [20].

False positives pose clinical risks including alarm fatigue and unnecessary interventions. Responsible deployment requires minimising false alarms through careful thresholding, using clinician-review escalation pathways, and monitoring

models for performance drift. A human-in-the-loop approach is essential, with automated alerts supporting rather than replacing clinical judgement. Alarm-management research shows that successful deployment must align with clinical workflows and involve continuous evaluation [11].

#### E. Limitations

Despite promising results, several limitations must be acknowledged. First, the dataset used in this study is simulated rather than real clinical data, which limits the generalisability of the findings. Real world physiological data often contain greater variability, unexpected noise, and patient specific complexity not reflected in simulated environments [19]. Second, the absence of labelled ground truth data prevents evaluation of true anomaly accuracy. Without labels, it is not possible to determine whether detected anomalies correspond to actual medical events or harmless fluctuations, making it difficult to quantify performance metrics such as sensitivity or specificity. Third, the dataset includes only a narrow set of vital signs [20]. In real IoT healthcare systems, additional variables such as oxygen saturation, respiratory rate, movement data, and environmental context play a significant role in anomaly interpretation.

Finally, both models have inherent algorithmic constraints. DBSCAN is highly sensitive to parameter choices such as eps and min samples, which can drastically alter results across datasets [18], [8]. KNN depends heavily on scaling and may be affected by irrelevant or noisy features, as noted in prior studies on distance based learning [25]. These limitations indicate that model performance may vary significantly outside controlled experimental conditions.

## CONCLUSION

This study shows that combining distance- and density-based methods improves anomaly detection in IoT medical sensor streams. KNN captures point-level deviations in physiological data, while LOF enhances sensitivity to local density changes, detecting subtle fluctuations in patient vitals [27]. DBSCAN isolates structural anomalies and small clusters of rare, clinically relevant events without assuming a fixed cluster number, improving robustness against noise and sensor variability [27, 28].

Overall, the combined KNN/LOF + DBSCAN approach provided complementary anomaly detection signals point wise, density aware, and cluster level improving reliability compared with any single model. This aligns with prior evidence that hybrid anomaly detection frameworks enhance early event recognition in medical cyber physical systems and can reduce false alarms, which remain a major challenge in continuous patient monitoring [29]. While further testing on diverse sensor modalities and under real time constraints is still required, the results suggest that such hybrid unsupervised methods have strong potential for supporting trustworthy IoT based clinical monitoring pipelines.

## REFERENCES

- [1] A. Khan, M. Rizwan, O. Bagdasar, A. Alabdulatif, S. Alamro, and A. Alnajim, "Deep Learning Driven Anomaly Detection for IoMT Based Smart Healthcare Systems," Journal/Conference Name, Year.
- [2] J. G. Hannah, A. S. D. Murthy, G. Kalnoor, M. Vetrivelan, and M. S. Nidhya, "Machine Learning Algorithms for Anomaly Detection in IoT Networks," Journal/Conference Name, Year.
- [3] J. Glova, T. Sabol, and V. Vajda, "Business models for the internet of things environment," Procedia Economics and Finance, vol. 15, pp. 1122–1129, 2014.
- [4] M. N. Bhuiyan, M. M. Rahman, M. M. Billah, and D. Saha, "Internet of things (IoT): A review of its enabling technologies in healthcare applications, standards protocols, security, and market opportunities," IEEE Internet of Things Journal, vol. 8, no. 13, pp. 10474–10498, 2021.
- [5] M. Hrvnak, L. Chen, A. Dubrawski, E. Bose, L. Clermont, and M. R. Pinsky, "Real alerts and artifact classification in archived multi signal vital sign monitoring data: implications for mining big data," J. Clin. Monit. Comput., vol. 30, no. 6, pp. 875–888, 2016.
- [6] A. Dey et al., "Weakly supervised classification of vital sign alerts as real or artifact," 2023. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/3712838/>
- [7] U.S. Food and Drug Administration, "Ventilator recall Life2000 ventilator due to potential failure battery charging dongle," Jul. 2024. [Online]. Available: <https://www.fda.gov/medical-devices/medical-device-recalls-and-early-alerts/ventilator-recall-baxter-removes-life2000-ventilator-due-potential-failure-battery-charging-dongle>
- [8] K. Liberatore and B. Kohler, "Batteries gone bad: Batteries power countless medical devices reliable performance essential," Patient Safety, 2019.
- [9] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," IEEE Trans. Inf. Theory, vol. 13, no. 1, pp. 21–27, 1967.
- [10] B. Boehmke, "K Nearest Neighbors," in \*Hands On Machine Learning\*, 2019. [Online]. Available: <https://bradleyboehmke.github.io/HOML/knn.html>
- [11] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density based algorithm for discovering clusters in large spatial databases with noise," in \*Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining\*, 1996, pp. 226–231.
- [12] Scikit learn Developers, "DBSCAN scikit learn documentation," 2025. [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html>
- [13] C. Schröer et al., "A systematic literature review on applying CRISP DM," Procedia Comput. Sci., vol. 184, pp. 104–115, 2021.
- [14] I. A. Bozdog, T. Daniel Nicusor, M. Antal, C. Antal, T. Cioara, I. Anghel, and I. Salomie, "Human behavior and anomaly detection using machine learning and wearable sensors," in \*2021 IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP)\*, pp. 383–390, Oct. 2021.
- [15] G. Zachos, I. Essop, G. Mantas, K. Porfyrikis, J. C. Ribeiro, and J. Rodriguez, "An anomaly based intrusion detection system for internet of medical things networks," \*Electronics\*, vol. 10, no. 21, p. 2562, 2021.
- [16] O. F. Jeelani, M. Njie, and V. M. Korzhuk, "Advancing Intrusion Detection in IoT Enabled Healthcare System Using K Nearest Neighbor Algorithm," in \*2025 IEEE 13th International Conference on Smart Energy Grid Engineering (SEGE)\*, pp. 231–236, Aug. 2025.
- [17] S. Subha and J. G. R. Sathiaseelan, "Effective anomaly detection approach to classify noisy data using robust noise detection and removal technique in IoT healthcare data," \*SN Computer Science\*, vol. 4, no. 5, 2023.
- [18] Information Commissioner's Office (ICO), "Health data," 2024. [Online]. Available: [https://ico.org.uk/for-organisations/uk\\_gdpr/guidance-and-resources/individual-rights/right-of-access/health-data/](https://ico.org.uk/for-organisations/uk_gdpr/guidance-and-resources/individual-rights/right-of-access/health-data/)
- [19] N. Norori et al., "Addressing bias in big data and AI for health care," \*NPJ Digital Medicine\*, 2021. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8515002/>
- [20] R. J. Chen et al., "Algorithmic fairness in artificial intelligence for medicine and healthcare: a perspective," 2023. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/37380750/>
- [21] B. Ahamed et al., "Design of an energy efficient IoT device assisted wearable for healthcare data management," 2023.
- [22] J. W. Anderson et al., "Algorithmic individual fairness and healthcare: a scoping review," \*JAMIA Open\*, 2025.

- [23] Sunny, J.S., Patro, C.P.K., Karnani, K., Pingle, S.C., Lin, F., Anekoji, M., Jones, L.D., Kesari, S. and Ashili, S., 2022. Anomaly detection framework for wearables data: a perspective review on data concepts, data analysis algorithms and prospects. *Sensors*, 22(3), p.756.
- [24] Mohy Eddine, M., Guezzaz, A., Benkirane, S. and Azrour, M., 2023. An efficient network intrusion detection model for IoT security using K NN classifier and feature selection. *Multimedia Tools and Applications*, 82(15), pp.23615–23633.
- [25] Gaddam, A., Wilkin, T., Angelova, M. and Gaddam, J., 2020. Detecting sensor faults, anomalies and outliers in the internet of things: A survey on the challenges and solutions. *Electronics*, 9(3), p.511.
- [26] IBRAHIM ADAMU, A.B.D.U., Donta, P.K., MOHD ALI, D.A.R.M.A.W.A.T.Y., SARANG, S., M STOJANOVIC, G.O.R.A.N. and SEROJA SARNIN, S.U.Z.I., 2025. A Systematic Literature Review of Advanced Machine Learning Techniques in Wireless Body Area Networks: Application, Challenges, and Future Directions. *IEEE Access*.
- [27] M. Gupta, J. Gao, C. Aggarwal, and J. Han, “Outlier Detection for Temporal Data: A Survey,” *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 9, pp. 2250–2267, 2014.
- [28] N. A. Alrajeh, M. Bashir, and B. Shams, “A Secure and Efficient Clustering Protocol for Wireless Sensor Networks,” *Int. J. Distrib. Sens. Netw.*, vol. 11, no. 4, pp. 1–11, 2015. (Uses DBSCAN for identifying irregular sensor behavior in IoT health networks.)
- [29] R. Ramadan and M. Alkasassbeh, “An Anomaly Detection Algorithm for Real Time IoT Health Data,” *Procedia Comput. Sci.*, vol. 170, pp. 705–712, 2020.
- [30] A. R. Javed, S. Usman, S. U. Rehman, N. Chilamkurti, and M. A. Jan, “Anomaly Detection in IoT Systems Using Hybrid Machine Learning,” *Sensors*, vol. 20, no. 18, p. 5212, 2020.

## APPENDIX

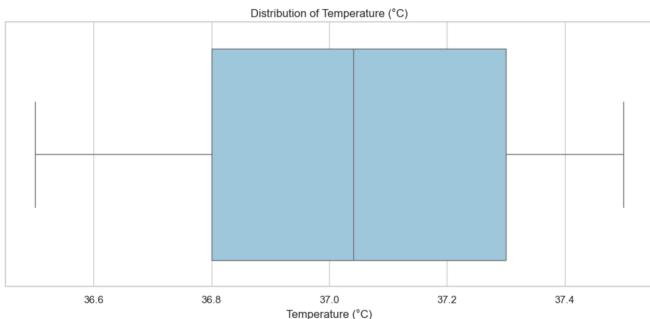


Fig. 6: Distribution of Temperature

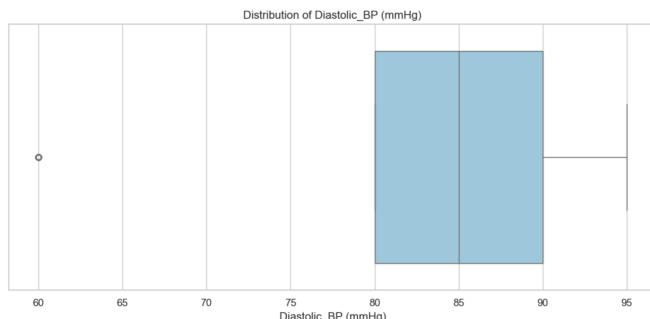


Fig. 7: Distribution of Diastolic BP

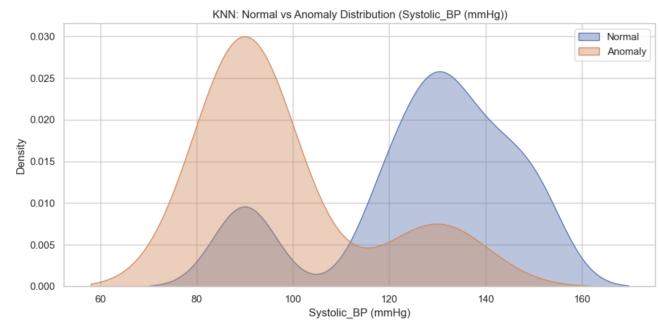


Fig. 9: KNN Normal vs Anomaly Systolic BP

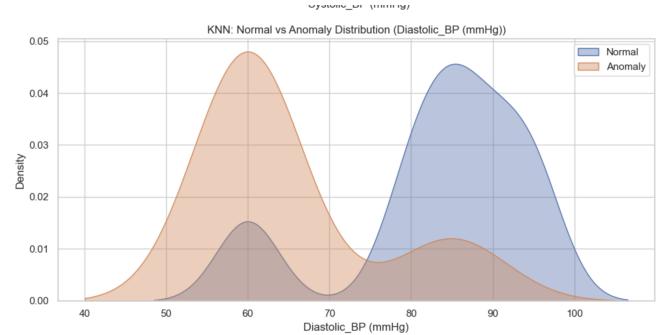


Fig. 10: KNN Normal vs Anomaly Diastolic BP

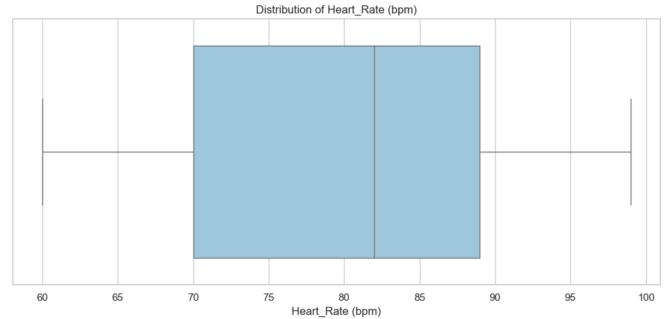


Fig. 8: Distribution of Heart Rate

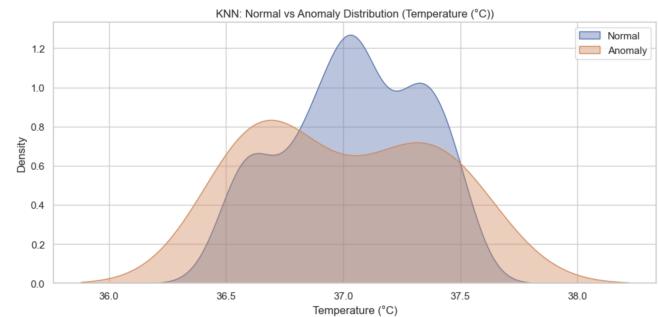


Fig. 11: KNN Normal vs Anomaly Temperature

Word limit excluding references = 3289 words