Anomaly detection in walking data using isolation forest: an unsupervised learning approach

Nur Alamsyah[1], Budiman[2], Valencia Claudia Jennifer Kaunang[3]
[1,2,3]Information System
Universitas Informatika Dan Bisnis Indonesia, Bandung, Indonesia
[1]nuralamsyah@unibi.ac.id, [2]budiman@unibi.ac.id, [3]valencia@unibi.ac.id

Submitted: January 18, 2025; accepted: October 30, 2025

Abstract

Detecting anomalies in walking data is crucial for ensuring data quality in wearable devices and understanding irregular physical activity patterns. Traditional methods often rely on labeled data, which is scarce in real-world applications. This study presents an unsupervised learning approach using Isolation Forest to detect anomalies in walking datasets. The data, comprising features such as step count, distance, and time, was preprocessed and analyzed to identify patterns and deviations. Isolation Forest was employed due to its efficiency in handling high-dimensional data and its ability to separate anomalies without prior labeling. The model successfully detected 5 anomalous data points out of the dataset, with anomaly scores ranging from -0.15 to 0.2. These outliers corresponded to extreme walking patterns, such as unusually high step counts with disproportionate time and distance. Visualization of anomaly scores and statistical evaluations validated the model's effectiveness, showing clear distinctions between normal and abnormal data. The proposed approach highlights the potential of Isolation Forest in improving data quality and enabling real-time anomaly detection in fitness tracking applications. This work contributes to the broader field of unsupervised anomaly detection by demonstrating a scalable and effective method for handling real-world activity data.

Keywords*: Anomaly Detection, Isolation Forest, Unsupervised Learning, Walking Data, Fitness Tracking.*

1. Introduction

The increasing use of wearable devices for tracking physical activities has led to an abundance of walking data, offering valuable insights into users' health and activity patterns [1]. However, these datasets often contain anomalies due to irregular user behavior, sensor errors, or data collection inconsistencies. Detecting and addressing such anomalies is crucial to ensure the reliability of fitness tracking applications and related studies [2]. Traditional methods for anomaly detection rely on supervised learning, which requires labeled data that is scarce and expensive to obtain, particularly in real-world applications [3]. While supervised approaches have shown success in various anomaly detection tasks, they rely heavily on labeled datasets, which are often scarce, imbalanced, or difficult to obtain—especially in real-world applications involving wearable data. Labeling anomalies in walking activity, for example, is highly subjective and context-dependent, making it challenging to define ground truth. Moreover, supervised models can be biased toward the majority class, leading to poor generalization when facing unseen or rare anomalies. In contrast, unsupervised learning approaches such as Isolation Forest do not require labeled data and are more robust in detecting novel or unexpected deviations. This makes them particularly suitable for anomaly detection in dynamic and noisy environments like wearable sensor data, where the types of anomalies can vary and evolve over time. This limitation has driven the need for unsupervised learning methods capable of identifying outliers without prior labeling [4]. Recent advancements in unsupervised learning techniques, such as Isolation Forest, have shown promise in detecting anomalies efficiently in high-dimensional datasets [5].

Isolation Forest, introduced by Liu *et al*. [6] isolates anomalies by constructing random decision trees and identifying data points that require fewer splits to separate. This method has demonstrated its efficacy across various domains, including fraud detection Tabrizchi *et al.* [7] sensor network monitoring devi *et al.* [8], and industrial applications aldrich *et al*. [9] Moreover, Isolation Forest is computationally efficient, making it suitable for large-scale datasets. In the context of physical activity data, anomaly detection has been employed to improve the reliability of wearable devices. Previous studies have applied varians approaches to anomaly detection in wearable and activity-related datasets. For instance, Khan *et al.* [10] employed supervised learning techniques for anomaly detection in IoT-based healthcare, requiring

extensive labeled data for model training, while zorriassatine *et al.* [11] focused on gait anomaly detection using infrared sensor arrays, emphasizing feature-specific analysis rather than general walking behavior. These studies highlight the growing importance of anomaly detection in ensuring the quality and usability of wearable data. In contrast, this study highlights the practicality and efficiency of the standard Isolation Forest algorithm applied to multidimensional walking data, aiming to fill the gap in research that focuses on simple yet effective unsupervised approaches for anomaly detection in wearable sensor data.

Despite these advancements, limited attention has been given to walking data that incorporates multiple features such as step count, distance, and walking time. These features collectively offer a more comprehensive understanding of walking behavior and its irregularities. To address this gap, this study leverages Isolation Forest as an unsupervised learning approach for detecting anomalies in walking datasets. By identifying outliers, this research aims to enhance the reliability and accuracy of walking data analysis, which is crucial for applications in healthcare, wearable technology, and fitness tracking. The objectives of this research are to detect anomalies in walking data using Isolation Forest, to evaluate the effectiveness of Isolation Forest in distinguishing normal and anomalous data, and to analyze the implications of anomaly detection for data quality improvement and real-world applications.

The contributions of this study include demonstrating the application of Isolation Forest for anomaly detection in walking data, providing a systematic evaluation of the model's performance through anomaly score distribution and visualizations, and highlighting the potential of unsupervised learning in addressing challenges in wearable data analysis and fitness tracking.

2. Method

The methodology used to detect anomalies in walking data is based on Isolation Forest, an unsupervised learning approach. The research framework is divided into four main stages: data collection, exploratory data analysis, anomaly detection modeling, and evaluation. Each stage is designed to systematically process and analyze the walking dataset to ensure the effective identification of anomalous data points. The proposed methodology is illustrated in Fig 1, emphasizing the importance of preprocessing, visualizing data patterns, applying a robust anomaly detection algorithm, and validating the results through statistical and visual evaluation. The following subsections provide detailed descriptions of each stage in the methodology.

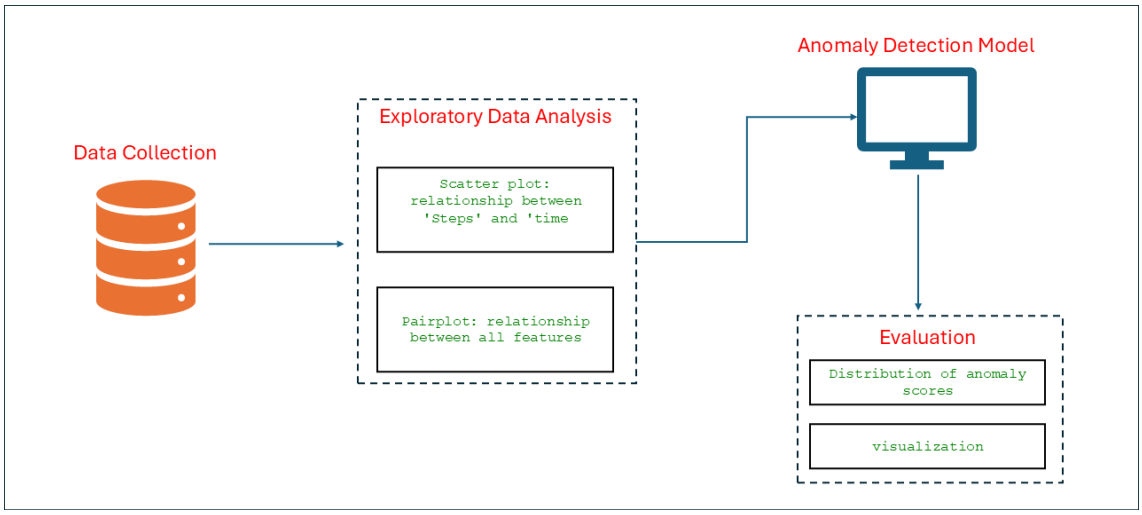

Fig. 1 Proposed method

2.1 Data Collection

The walking data used in this study was obtained from the Kaggle portal, a popular platform for data science and machine learning projects. This data is taken from the smartwatch. The dataset comprises 95 records, with each record representing a unique instance of walking activity. The dataset includes three main features: step count, distance, and walking time. These features are essential for analyzing walking behavior and detecting anomalies effectively. A summary of the dataset features is provided in Tabel 1.

Table 1. Data Description

Feature	Description	Unit	Example
Steps	Total number of steps taken	Steps	12345
km	Distance covered during the activity	Kilometers	7.5
time	Time spent walking	Minutes	120

2.2 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was performed to examine the relationships between features and detect potential patterns or irregularities in the walking dataset [12]. A scatter plot was created to visualize the relationship between Steps and time, with km represented as the hue and size of the points. This visualization provided a detailed view of how walking distance correlates with step count and walking time, highlighting any deviations from expected trends. The use of a color gradient and varying point sizes effectively emphasized the role of distance in these interactions [13].

Additionally, a pairplot was generated to analyze the relationships between all features, including Steps, km, and time. This pairplot included scatter plots for feature combinations and kernel density estimates (KDE) for individual feature distributions, offering a comprehensive view of feature interdependencies and distribution patterns. Together, these visualizations allowed for the identification of potential anomalies or irregular data points that could influence the performance of the anomaly detection model. These EDA steps ensured a deeper understanding of the dataset's structure and informed the design of the subsequent anomaly detection phase.

2.3 Anomaly Detection Model

Anomaly detection in this study was performed using the Isolation Forest algorithm, a robust and computationally efficient unsupervised learning method. Isolation Forest identifies anomalies by isolating data points that deviate significantly from the majority. It constructs random decision trees and measures the path length required to isolate each data point. Points with shorter path lengths are more likely to be anomalies.

The model was implemented using the walking dataset features ‘Steps, ‘km’, and ‘time’. The contamination parameter was set to 0.05, assuming that 5% of the data are anomalies. This parameter determines the proportion of data points expected to be anomalous and directly affects the model’s sensitivity. After fitting the model, an anomaly score was calculated for each data point, and data points with scores indicating significant deviation were labeled as anomalies. The anomaly score $(s(x))$ for a data point (x) is computed using the following equation:

$$s(x) = 2^{-\frac{E(h(x))}{c(n)}} \quad (1)$$

Where$(h(x))$: Path length of data point (x), $(E(h(x)))$: Average path length of (x) across all trees in the forest and $(c(n))$: Average path length of an unsuccessful search in a binary tree, approximated as:

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n} \quad (2)$$

With $(H(i))$ being the (i)-th harmonic number, calculated as:

$$H(i) = \sum_{k=1}^{i} \frac{1}{k} \quad (3)$$

Data points with scores closer to 1 are considered normal, while those with scores closer to 0 are identified as anomalies. The model output includes binary labels for each data point, with -1 indicating an anomaly and 1 indicating normal data. A scatter plot was generated to visualize the detected anomalies. In the plot, data points classified as anomalies are highlighted in red, while normal data points are shown in blue. The size of the points corresponds to the distance ‘km’, adding an additional dimension to the visualization.

2.4 Evaluation

The evaluation of the Isolation Forest model was conducted by analyzing the anomaly scores and visualizing their impact on the dataset. Anomaly scores, which represent the degree of deviation of each data point from the majority, were computed for all data points in the dataset. Lower scores indicated a higher likelihood of being an anomaly, whereas higher scores corresponded to normal data points. The distribution of anomaly scores was visualized using a histogram, which revealed that the majority of data points had scores near the upper end of the range, signifying normal behavior. A small portion of the data exhibited significantly lower scores, aligning with the points identified as anomalies. The histogram included a kernel density estimation (KDE) overlay, providing a smooth representation of the distribution and helping to highlight the separation between normal and anomalous data [14].

To gain further insights, the anomaly scores were visualized on a scatter plot. The scatter plot mapped the relationship between Steps and time, with the color of each point reflecting its anomaly score. Points with higher anomaly scores, representing normal data, were displayed in shades of blue, while points with lower scores, indicating anomalies, appeared in shades of red. Additionally, the size of each point was proportional to the walking distance (km), adding another layer of information to the visualization. This scatter plot highlighted how anomalies deviated from typical walking patterns, often clustering away from normal data points. For instance, points with unusually low anomaly scores stood out visually, confirming the model's ability to detect significant deviations [15].

Overall, the evaluation demonstrated that the Isolation Forest model effectively distinguished normal data points from anomalies. The histogram confirmed the model's separation of data into distinct score ranges, while the scatter plot provided an intuitive representation of the anomalies and their relationships with other features. These evaluations validated the model's performance and highlighted its capability to identify irregular walking behaviors within the dataset.

3. Result and Discussion

The findings of the study are organized into two parts: Exploratory Data Analysis (EDA), anomaly detection using the Isolation Forest model, and the evaluation of model performance. These results provide insights into the dataset's structure, the anomalies identified, and the model's ability to separate anomalous data points from normal ones. Furthermore, a discussion is included to compare the findings of this study with existing literature, highlighting the strengths and limitations of the proposed approach. This comprehensive analysis demonstrates the potential of Isolation Forest as an unsupervised learning method for detecting anomalies in walking data.

3.1. Result

Fig. 2 illustrates the relationship between the number of steps taken (Steps) and the time spent walking (Time) in minutes, with the walking distance (km) represented as the hue and size of the data points. The plot demonstrates a clear positive correlation between Steps and Time, indicating that as the number of steps increases, the walking time also tends to increase.

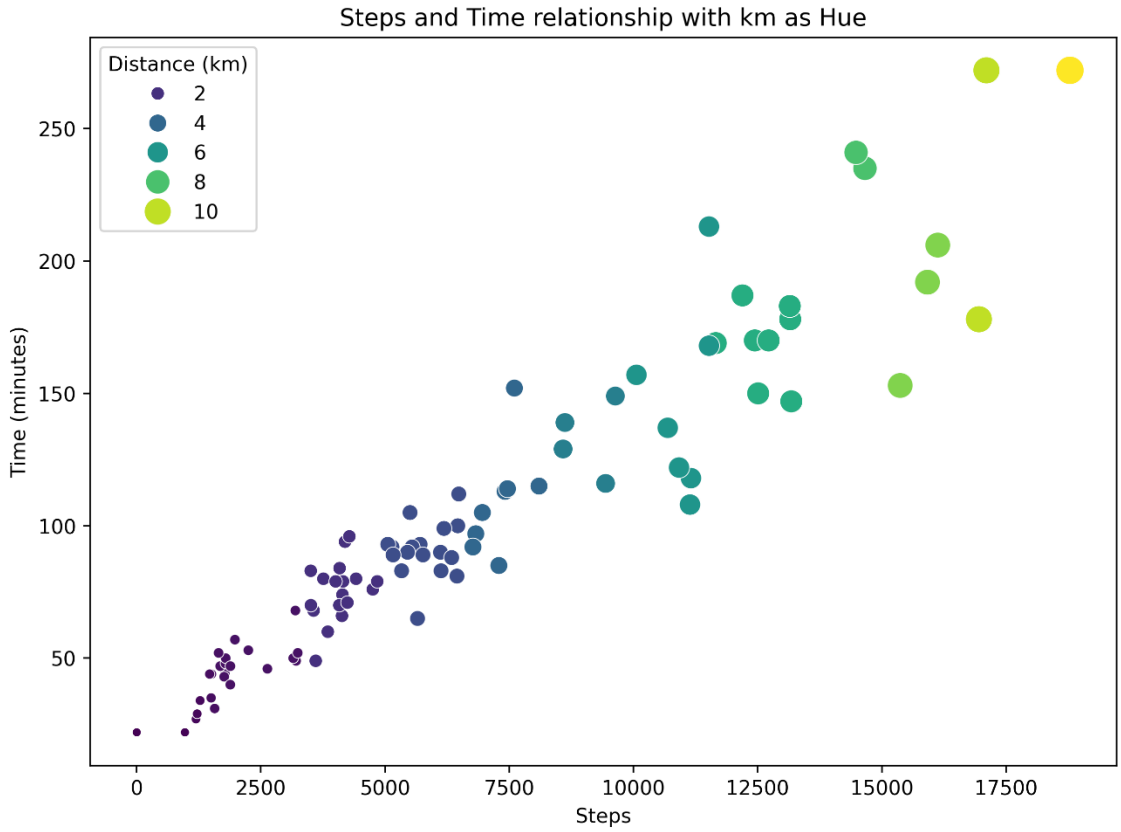

Fig. 2 Relationship between the number of steps taken

The color gradient and varying point sizes highlight the role of walking distance (km) in this relationship. Shorter distances, represented by smaller points in darker colors (e.g., purple and blue), cluster in the lower left of the plot, where both Steps and Time are relatively low. Conversely, longer distances, represented by larger points in lighter colors (e.g., green and yellow), are distributed towards the upper right, corresponding to higher Steps and Time values. This visualization provides valuable insights into the dataset's structure, revealing proportional relationships between the features and helping to identify potential anomalies. For instance, data points that deviate significantly from the general trend, such as those with unusually high Time values for relatively few Steps, may warrant further investigation as potential anomalies.

Fig. 3 presents a pairplot that visualizes the relationships between the features in the walking dataset, including Steps, km, time, and additional attributes generated during anomaly detection, such as anomaly_score, is_anomaly, and anomaly_score_value. The diagonal plots represent the distribution of individual features using kernel density estimation (KDE), while the off-diagonal scatter plots depict the pairwise relationships between the features. A strong positive linear relationship is observed between Steps and km, as higher step counts are naturally associated with greater distances covered. Similarly, a positive correlation exists between Steps and time as well as km and time, indicating that longer walking durations correspond to higher step counts and longer distances.

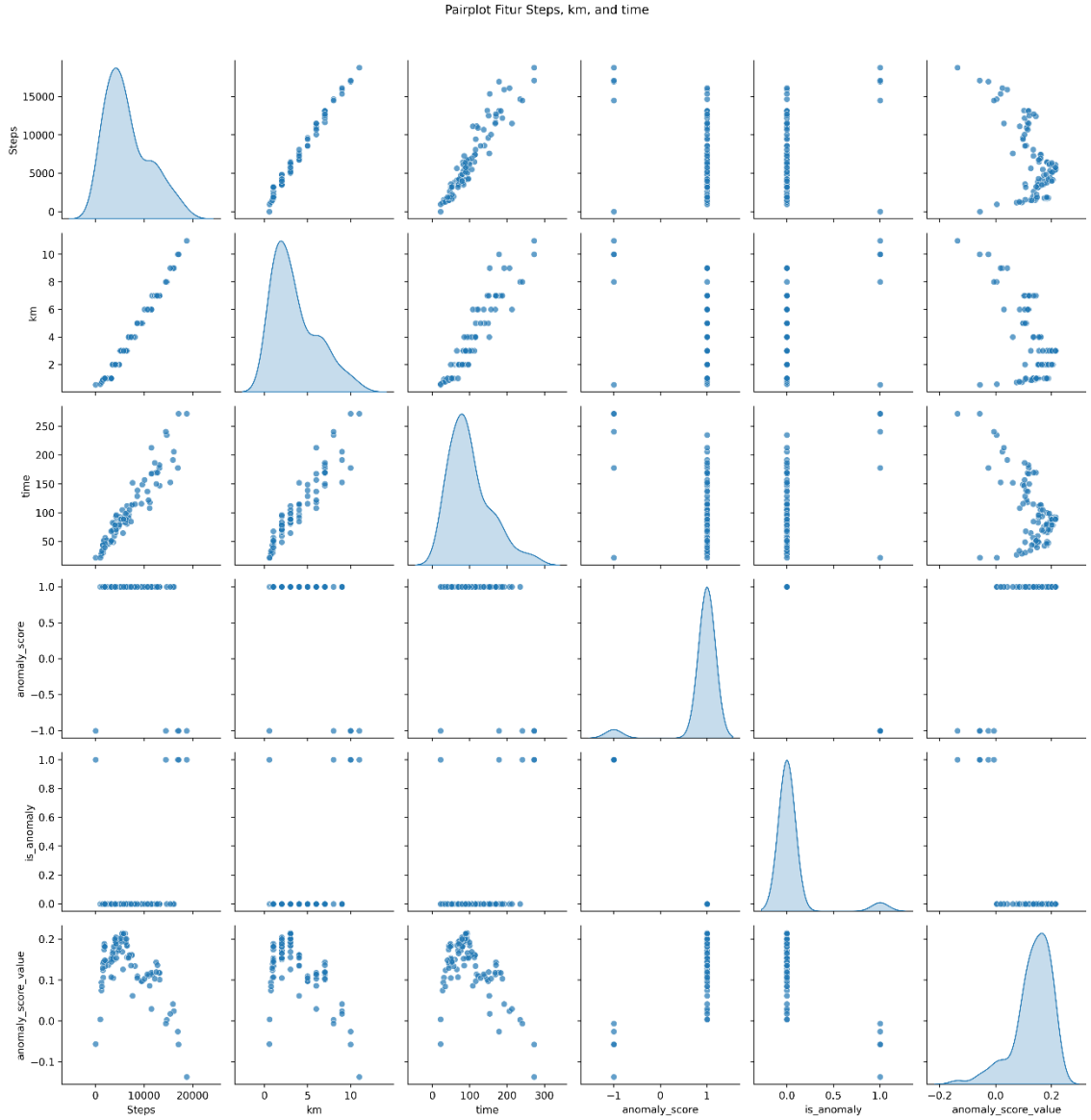

Fig. 3 Relationships between the features

The anomaly-related features provide additional insights into the data. The anomaly_score feature helps identify points with significant deviations, where lower scores indicate a higher likelihood of being anomalies. The binary feature is_anomaly clearly differentiates normal data points (labeled as 0) from anomalies (labeled as 1), with anomalous points visibly scattered away from the primary clusters in certain plots. Furthermore, the anomaly_score_value feature, which represents a continuous measure of anomaly severity, highlights deviations across various feature combinations, with values closer to 0 indicating significant outliers. This pairplot offers a comprehensive overview of feature interdependencies while integrating anomaly detection results, enabling the identification of patterns, relationships, and irregularities within the dataset.

Table 2 presents the results of anomaly detection using the Isolation Forest model, where five data points were identified as anomalies. The table includes features such as Steps, km, and time, along with the anomaly scores (anomaly_score) and labels (is_anomaly) assigned by the model. Anomalous data points are marked with a label of -1 in the anomaly_score column and 1 in the is_anomaly column. For example, the first row shows a data point with only 0.9 steps (Steps) recorded for a distance of 0.54 km in 22 minutes. The anomaly score for this data point is very low, indicating significant deviation from the rest of the dataset.

Table 2. Example of Anomalous Data Detected

Steps	km	time (minutes)	anomaly_score	is_anomaly	anomaly_score_value
0.9	0.54	22	-1	1	-0.055884
16955.0	10.00	178	-1	1	-0.025653
18788.0	11.00	272	-1	1	-0.136597
14482.0	8.00	241	-1	1	-0.005958
17106.0	10.00	272	-1	1	-0.057107

Fig. 4 provides a detailed visualization of the anomaly detection results using a scatter plot to represent the relationship between Steps (x-axis) and time (y-axis). In this plot, anomalous data points are highlighted in red, while normal data points are shown in blue, making it easy to distinguish between the two groups. The size of the points also corresponds to the distance covered (km), adding an additional layer of interpretability to the visualization.

The plot reveals that anomalies tend to lie outside the general linear pattern observed between Steps and time, where an increase in step count generally corresponds to a proportional increase in walking time. These anomalies could represent irregularities, such as data points with unusually high or low step counts relative to the time spent walking. For instance, a data point with an extremely low step count but a relatively long walking time, or vice versa, stands out as a deviation from the main cluster of data. Such deviations may arise due to errors in data collection, user behavior, or other external factors influencing the walking activity.

This visualization underscores the effectiveness of the Isolation Forest model in isolating and identifying outliers. By assigning an anomaly score to each data point, the model successfully detected those that differ significantly from the overall trend, providing valuable insights into the dataset. The clear separation between normal and anomalous data points in Fig. 4 highlights the robustness of the model in handling multidimensional data and identifying patterns of irregular walking behavior.

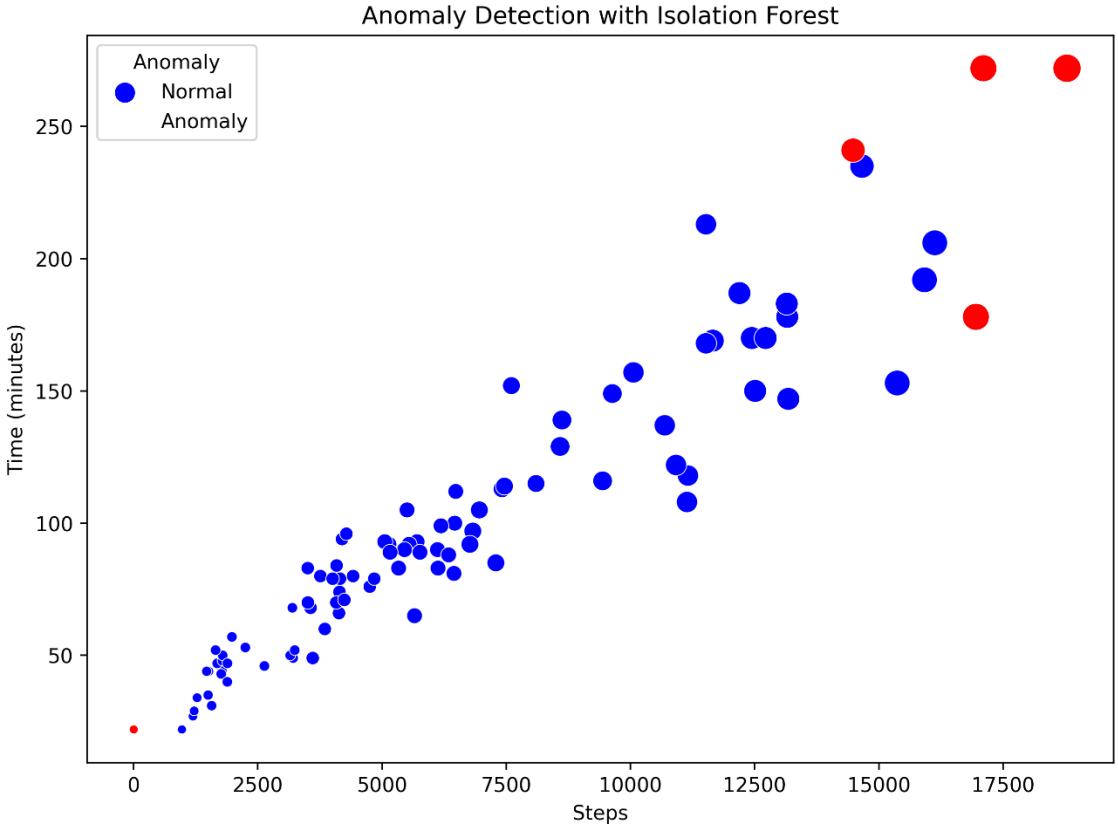

Fig. 4 Anomaly detection results

Fig. 5 illustrates the distribution of anomaly scores assigned by the Isolation Forest model for the walking dataset. The x-axis represents the anomaly score, a numerical value indicating the degree of

deviation of each data point from the general patterns observed in the dataset. A higher score (closer to 0.2) suggests that the data point is normal and consistent with the overall dataset trends, while lower scores (closer to -0.1 or below) indicate a higher likelihood of being an anomaly. The y-axis represents the frequency of data points within specific anomaly score ranges, making it possible to observe the overall distribution of the dataset based on the model's scoring.

The histogram, overlaid with a Kernel Density Estimation (KDE) curve, provides a detailed view of how the data is distributed across different anomaly scores. The majority of the data points cluster around higher anomaly scores (above 0.1), suggesting that most of the dataset aligns well with the expected trends and is classified as normal. In contrast, a smaller number of data points exhibit significantly lower anomaly scores (below 0), representing potential anomalies that deviate substantially from the general data distribution.

This visualization demonstrates the effectiveness of the Isolation Forest model in distinguishing normal data points from anomalous ones. The clear separation between the higher anomaly scores (normal data) and the lower anomaly scores (anomalies) highlights the model's capability to assign scores based on the degree of isolation of each data point. The KDE curve smoothens the score distribution, emphasizing the concentration of normal data and the sparse, isolated nature of anomalies. The peaks in the KDE curve correspond to the densest regions of normal data, while the troughs highlight areas where anomalous points are located. Furthermore, this evaluation validates the assumption that anomalies constitute a small proportion of the dataset, as reflected in the limited frequency of low anomaly scores. The model's ability to assign scores consistently across the dataset ensures that even subtle deviations in the data can be captured effectively.

The insights from Fig. 5 reinforce the Isolation Forest model's suitability for anomaly detection tasks in multidimensional datasets, such as walking data, by leveraging the relationships among features like Steps, km, and time. The visualization also serves as a diagnostic tool, confirming the model's ability to separate normal and anomalous data in a clear and interpretable manner.

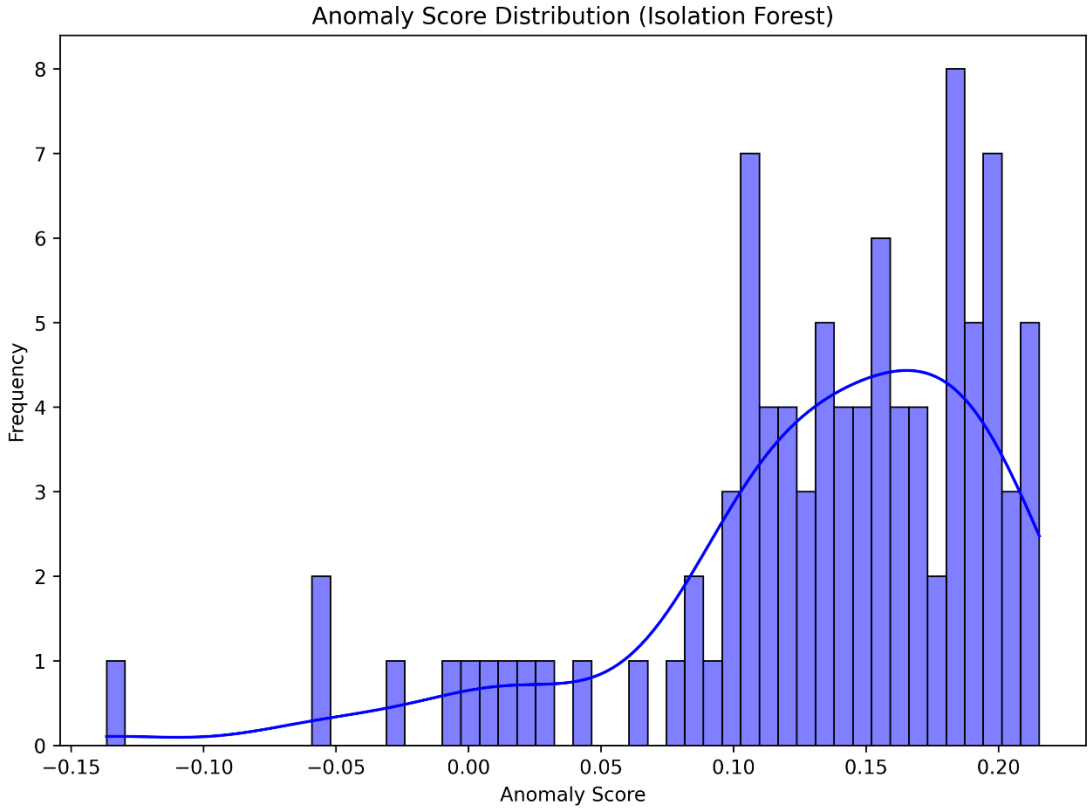

Fig. 5 Distribution of anomaly scores

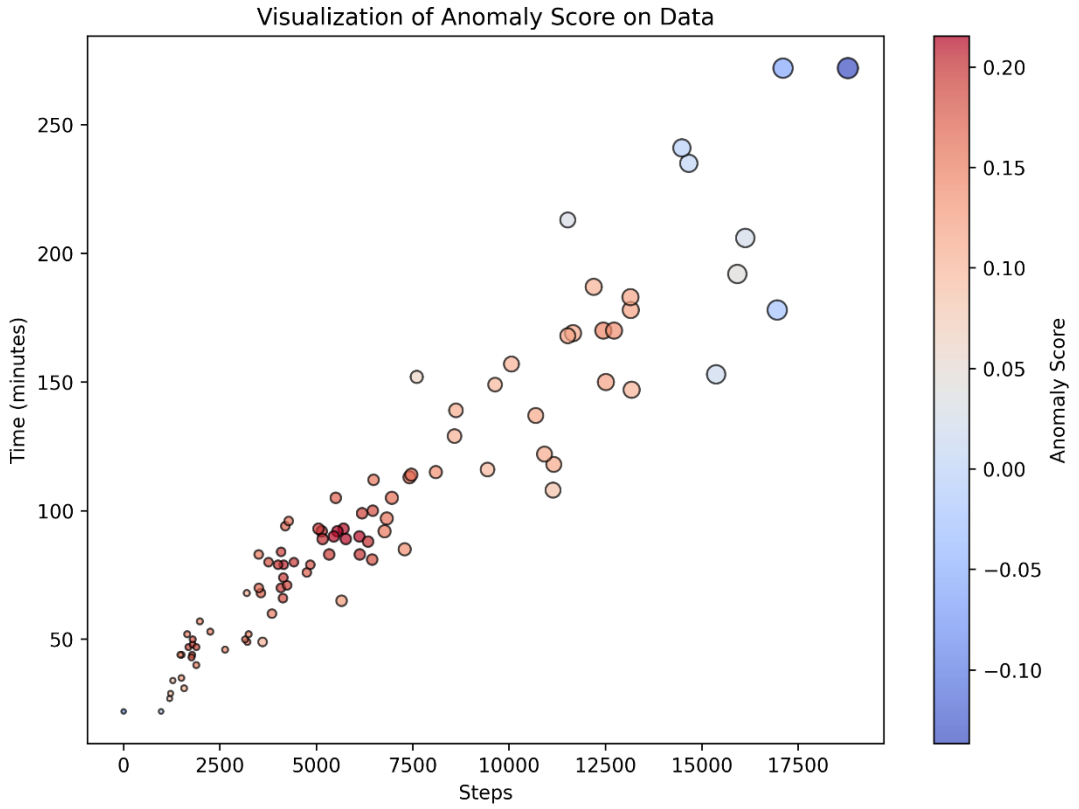

Fig. 6 Distribution of anomaly scores

Fig. 6 visualizes the anomaly scores generated by the Isolation Forest model on the walking dataset. The scatter plot displays the relationship between Steps (x-axis) and time (y-axis), with the anomaly score represented by the color gradient of the points. The colorbar on the right indicates the range of anomaly scores, where darker shades of red correspond to higher anomaly scores (closer to 0.2, indicating normal data), and darker shades of blue represent lower anomaly scores (closer to -0.1, indicating anomalies). Additionally, the size of each point is proportional to the walking distance (km), adding another dimension to the analysis.

From the visualization, it is evident that most data points with higher anomaly scores (red shades) align well with the general linear trend between Steps and time, indicating that they are classified as normal data. Conversely, points with lower anomaly scores (blue shades) deviate from this trend and are potential anomalies. For example, data points with disproportionately low or high step counts relative to the time spent walking are assigned lower anomaly scores, suggesting significant deviations from expected patterns.

The scatter plot effectively highlights the distribution of anomaly scores across the dataset and provides insights into how the Isolation Forest model evaluates data points based on their degree of deviation. The smooth gradient of colors reflects the model's sensitivity in scoring data points, capturing both minor and significant deviations. This visualization complements other evaluation metrics by illustrating the spatial distribution of anomaly scores and their alignment with feature relationships. It reinforces the effectiveness of the Isolation Forest model in detecting outliers and showcases its ability to assign anomaly scores that correlate with deviations in multidimensional data.

3.2. Discussion

We compare our findings with those of previous studies to contextualize the effectiveness of the Isolation Forest model in anomaly detection within walking datasets. Liu et al. [6] introduced the Isolation Forest algorithm, emphasizing its efficiency in detecting anomalies by isolating them through recursive partitioning. Their work demonstrated the model's capability to handle high-dimensional data and its computational efficiency, which aligns with our findings in detecting anomalies in walking data.

Xu *et al.* [16] proposed the Deep Isolation Forest, which enhances the traditional Isolation Forest by utilizing casually initialized neural networks to map original data into random representation ensembles. This approach facilitates high freedom of partitioning in the original data space, encouraging a unique synergy between random representations and random partition-based isolation. Their extensive experiments

showed significant improvement over state-of-the-art isolation-based methods and deep detectors on various datasets, supporting the potential for further enhancements in our approach.

Our study makes a significant contribution to the existing body of knowledge by focusing on the application of the Isolation Forest model to walking datasets, a domain where anomaly detection has practical importance but remains relatively underexplored. By leveraging the relationships among key features such as step count, walking distance, and time spent walking, the study highlights the model's capability to effectively identify irregular walking behaviors that deviate from established patterns. This application not only validates the Isolation Forest model's utility in handling multidimensional data but also underscores its adaptability to scenarios where labeled anomalies are scarce.

To strengthen the contextualization of our study, Table 3 presents a structured comparison between our research and several relevant prior works. It contrasts methods, data characteristics, label dependencies, and domain focus, while also identifying the key limitations of each approach. This comparative analysis underscores the novelty of applying a standard Isolation Forest model to real-world walking data for anomaly detection without requiring labeled datasets.

Table 3. Comparative Summary of Related Studies on Anomaly Detection

Study	Method	Data Type	Label Requirement	Application Domain	Key Limitation
Liu et al. (2008) [6]	Isolation Forest	General tabular datasets	Not required	Generic anomaly detection	Did not address wearable or activity data
Tabrizchi & Razmara (2024) [7]	Isolation Forest + GWO	Credit card transaction data	Not required	Fraud detection	Domain-specific, not applicable to physical activity
Khan & Alkhathami (2024) [10]	Supervised ML	IoT-based healthcare (step count)	Required	Healthcare anomaly detection	Needs large labeled data, not scalable
Zorriassatine et al. (2024) [11]	Supervised ML (sensor array)	Gait data (infrared sensors)	Required	Patient rehabilitation	Sensor-specific, lacks generalizability
Xu et al. (2023) [16]	Deep Isolation Forest	Benchmark anomaly datasets	Not required	Various anomaly datasets	Complex model with high computation cost
This Study	Isolation Forest (standard)	Walking data (steps, km, time)	Not required	Wearable / fitness tracking	Focused on real-world simplicity and interpretability

Furthermore, the alignment of our findings with previous research demonstrates the versatility of the Isolation Forest algorithm across various anomaly detection tasks. Its efficiency in isolating outliers and assigning anomaly scores makes it a robust choice for domains with complex feature interactions, such as walking data. This study reinforces the model's effectiveness, extending its applicability to new contexts while providing insights into potential real-world applications, such as healthcare monitoring, fitness tracking, and behavioral analysis. By building on the foundation established in prior research, our work bridges the gap between theoretical advancements and practical implementations, paving the way for future studies to explore its broader applications.

4. Conclusion

In this study, we successfully applied the Isolation Forest model to detect anomalies in walking datasets, focusing on features such as step count, walking distance, and time. The results demonstrate the model's effectiveness in identifying irregular walking behaviors that deviate from established patterns, with clear separation between normal data and anomalies. Through exploratory data analysis and model evaluation, we validated the robustness of the Isolation Forest in assigning anomaly scores and isolating outliers, even in multidimensional datasets. Compared to previous studies, our findings align with the demonstrated versatility and efficiency of the Isolation Forest algorithm in various anomaly detection tasks. This research contributes to the growing body of knowledge by showcasing its applicability in walking data, providing insights for real-world applications such as healthcare monitoring, fitness tracking, and behavioral analysis. Future work could explore integrating advanced variants of Isolation Forest or combining it with other machine learning techniques to further enhance anomaly detection performance.

References

[1] J. Qi, P. Yang, A. Waraich, Z. Deng, Y. Zhao, and Y. Yang, "Examining sensor-based physical activity recognition and monitoring for healthcare using Internet of Things: A systematic review," *J. Biomed. Inform.*, vol. 87, pp. 138–153, 2018.

[2] H. Banaee, M. U. Ahmed, and A. Loutfi, "Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges," *Sensors*, vol. 13, no. 12, pp. 17472–17500, 2013.
[3] S. Wang, J. F. Balarezo, S. Kandeepan, A. Al-Hourani, K. G. Chavez, and B. Rubinstein, "Machine learning in network anomaly detection: A survey," *IEEE Access*, vol. 9, pp. 152379–152396, 2021.
[4] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, and K. J. Kim, "A survey of deep learning-based network anomaly detection," *Clust. Comput.*, vol. 22, pp. 949–961, 2019.
[5] M. T. R. Laskar *et al.*, "Extending isolation forest for anomaly detection in big data via K-means," *ACM Trans. Cyber-Phys. Syst. TCPS*, vol. 5, no. 4, pp. 1–26, 2021.
[6] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 eighth ieee international conference on data mining*, IEEE, 2008, pp. 413–422.
[7] H. Tabrizchi and J. Razmara, "Credit card fraud detection using hybridization of isolation forest with grey wolf optimizer algorithm," *Soft Comput.*, pp. 1–19, 2024.
[8] S. K. Devi, R. Thenmozhi, and D. S. Kumar, "Self-Healing IoT Sensor Networks with Isolation Forest Algorithm for Autonomous Fault Detection and Recovery," in *2024 International Conference on Automation and Computation (AUTOCOM)*, IEEE, 2024, pp. 451–456.
[9] C. Aldrich and X. Liu, "Monitoring of Mineral Processing Operations with Isolation Forests," *Minerals*, vol. 14, no. 1, p. 76, 2024.
[10] M. M. Khan and M. Alkhathami, "Anomaly detection in IoT-based healthcare: machine learning for enhanced security," *Sci. Rep.*, vol. 14, no. 1, p. 5872, 2024.
[11] F. Zorriassatine, A. Naser, and A. Lotfi, "Gait Anomaly Detection with Low Cost and Low Resolution Infrared Sensor Arrays," *IEEE Sens. J.*, 2024.
[12] N. Alamsyah, B. Budiman, T. P. Yoga, and R. Y. R. Alamsyah, "XGBOOST HYPERPARAMETER OPTIMIZATION USING RANDOMIZEDSEARCHCV FOR ACCURATE FOREST FIRE DROUGHT CONDITION PREDICTION," *J. Pilar Nusa Mandiri*, vol. 20, no. 2, pp. 103–110, 2024.
[13] N. Alamsyah, B. Budiman, T. P. Yoga, and R. Y. R. Alamsyah, "COMPARISON LINEAR REGRESSION AND RANDOM FOREST MODELS FOR PREDICTION OF UNDERGROUND DROUGHT LEVELS IN FOREST FIRES," *J. Techno Nusa Mandiri*, vol. 21, no. 2, pp. 81–86, 2024.
[14] A. G. Putrada, I. D. Oktaviani, M. N. Fauzan, and N. Alamsyah, "CNN-LSTM for MFCC-based Speech Recognition on Smart Mirrors for Edge Computing Command," *J. Dinda Data Sci. Inf. Technol. Data Anal.*, vol. 4, no. 2, pp. 63–74, 2024.
[15] N. Alamsyah, A. P. Kurniati, and others, "A Novel Airfare Dataset To Predict Travel Agent Profits Based On Dynamic Pricing," in *2023 11th International Conference on Information and Communication Technology (ICoICT)*, IEEE, 2023, pp. 575–581.
[16] H. Xu, G. Pang, Y. Wang, and Y. Wang, "Deep isolation forest for anomaly detection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12591–12604, 2023.