Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 788-797

AI-Based Health Anomaly Detection Using Wearable Fitness Data

[1]Rashmita Das, [2]Pawan Jaiswal
[1]Student, [2]Assistant Professor
[1,2]Amity University Chhattisgarh

ABSTRACT

While wearable technology now makes it possible to track heart rate, activity levels, and sleep cycles around the clock, the sheer volume of data produced can be overwhelming and often lacks clear utility for the average user. This study introduces an AI-driven framework aimed at bridging that gap by detecting anomalies within wearable fitness datasets and translating them into actionable health insights. By processing structured device outputs through a specialized pipeline, the system employs a hybrid approach—integrating traditional statistical methods like Z-score analysis with machine learning models such as Isolation Forest—to pinpoint physiological irregularities. To ensure data integrity, our preprocessing stage standardizes disparate column formats, manages missing entries, and consolidates raw metrics into coherent daily summaries. These findings are then surfaced through an interactive dashboard, designed to help users navigate their health trends with greater clarity. Ultimately, by prioritizing a scalable and straightforward architecture, this approach offers a reliable means of personal health oversight and early risk identification, showing that a multi-layered detection strategy significantly strengthens the accuracy of identifying deviations from a user's baseline.

Keywords: Wearable Data, Health Monitoring, Anomaly Detection, Machine Learning, Z-score, Isolation Forest, FastAPI, Streamlit, Data Analytics, Fitness Tracking

I. INTRODUCTION

In recent years, the ubiquity of smartwatches and fitness bands has fundamentally changed how we track personal health, allowing for the constant recording of vitals like heart rate and sleep quality. While the resulting time-series data offers a wealth of information for preventive care, most current systems fall short of their potential. They typically provide rudimentary visualizations that fail to offer the sophisticated, automated interpretation needed to identify burgeoning health risks. The central challenge lies in converting these vast, raw data streams into the kind of intuitive intelligence that helps a person truly grasp their physiological state. Sudden irregularities, perhaps a sharp spike in resting heart rate or a series of disrupted sleep cycles, can be early warning signs of illness or stress, yet they often go unnoticed without specialized analytical support. To bridge this gap, researchers have increasingly turned to anomaly detection. Traditional statistical techniques are effective at flagging obvious outliers, whereas machine learning models are better suited to uncovering more nuanced, multivariate patterns. However, many existing solutions rely on a single methodology, which can be easily compromised by the inherent "noise" and variability of real-world wearable data. This paper details an AI-based detection system that improves reliability by combining multiple analytical approaches, including statistical modeling and behavioral pattern recognition. By refining raw fitness datasets through a dedicated preprocessing and feature engineering pipeline, the system identifies anomalies with greater precision. The findings are then presented through an interactive dashboard designed for clarity, aiming to move beyond simple data collection toward a more practical, intelligent solution for personal health monitoring.

II. LITERATURE REVIEW

Significant academic research has explored wearable health data analysis, driven by the increasing ubiquity of devices like smartwatches and fitness trackers. Traditional methodologies for identifying health deviations frequently rely on statistical instruments such as Z-score analysis and threshold-based evaluation. While effective at flagging extreme outliers by measuring standard deviations from a population or individual mean, these methods often struggle to capture the complex, multivariate relationships inherent in synchronized physiological signals like heart rate and body temperature.

To overcome the limitations of rigid statistical thresholds, recent studies have pivoted toward machine learning (ML), particularly unsupervised algorithms such as Isolation Forest. By utilizing recursive partitioning to isolate sparse data points, these models can identify anomalies in multidimensional datasets without requiring labeled training data—a critical advantage given that "ground truth" labels for health crises are rarely available in consumer wearable streams. Furthermore, deep learning architectures, such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN), have demonstrated superior performance over conventional methods like k-means clustering and Z-scores by modeling the temporal continuity and long-range dependencies of time-series data.

Personalization has emerged as another core focus, with researchers employing behavioral pattern analysis to compare current physiological data against a user's unique historical baseline. This shift recognizes that "normal" health metrics vary significantly across individuals, making population-level thresholds less reliable than person-centric trends. Time-series forecasting models, such as Facebook Prophet, have also been explored to predict expected health values and flag deviations, though their computational demands can complicate real-time edge deployment.

Despite these advancements, many existing systems remain limited by a reliance on single-modality detection, which often fails when confronted with the "noisy" or missing data typical of real-world wearable use. Emerging frameworks now emphasize hybrid integration, combining statistical rigor, unsupervised machine learning, and behavioral modeling to improve both accuracy and reliability in diverse healthcare scenarios.

III. THE PROPOSED SYSTEM

The proposed architecture is designed to evaluate wearable fitness data through a hybrid detection framework, merging statistical rigor with machine learning and behavioral modeling. The system ingests structured datasets encompassing core metrics such as heart rate, step count, and sleep duration. To ensure high data integrity, the pipeline begins with a preprocessing stage that standardizes column nomenclature, resolves missing values, and ensures consistent

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 788-797

formatting across all inputs. Following this, feature engineering is utilized to derive rolling averages and variability measures, providing a nuanced view of short-term trends and individual behavioral baselines.

At the center of the system is a multi-layered anomaly detection module. Extreme physiological deviations are identified using a Z-score approach, while the Isolation Forest model is employed to detect more complex, multivariate anomalies across disparate features. Furthermore, the system incorporates behavioral pattern analysis, comparing real-time values against historical trends to pinpoint personalized irregularities.

While the framework includes a time-series modeling component based on Facebook Prophet to capture temporal shifts and forecast expected values, its deployment is selective to manage computational overhead. To make these technical findings accessible, the system synthesizes all detected anomalies and trends into an interactive dashboard, allowing users to intuitively visualize their health data and recognize significant irregularities as they occur.

IV. METHODOLOGY

The proposed system operates through a structured pipeline designed to transform raw wearable fitness data into interpretable health intelligence. The methodology is organized into several distinct operational stages:

- **Data Ingestion**: The process begins by ingesting structured datasets comprising physiological and behavioral parameters, including heart rate, step count, sleep duration, and corresponding timestamps, which serve as the primary output from the wearable devices.
- **Data Preprocessing**: To ensure high analytical quality, the raw input undergoes a cleaning phase where column nomenclature is standardized, missing or invalid entries are resolved, and data types are converted for consistency. This stage is critical for neutralizing inconsistencies that might otherwise skew detection results.
- **Feature Engineering**: Once cleaned, the data is enriched through feature engineering to highlight underlying patterns. By calculating rolling statistics—such as moving averages and standard deviations over specific time windows—the system effectively captures short-term trends and subtle variations in a user’s daily behavior.
- **Anomaly Detection**: The core of the pipeline utilizes a hybrid detection strategy. A statistical Z-score method identifies extreme deviations in individual metrics, while an Isolation Forest model evaluates multivariate relationships to find anomalies that might be missed by looking at a single variable in isolation. Furthermore, behavioral analysis compares current metrics against historical baselines to identify personalized deviations. For deeper temporal insights, a Prophet time-series model can be optionally deployed to forecast expected values and flag unexpected shifts.
- **Visualization and Interpretation**: In the final stage, the analyzed data is rendered into visual formats, such as interactive graphs and trend lines. This graphical representation allows for a clear interpretation of health patterns, making it easier for users to recognize and act upon detected irregularities.

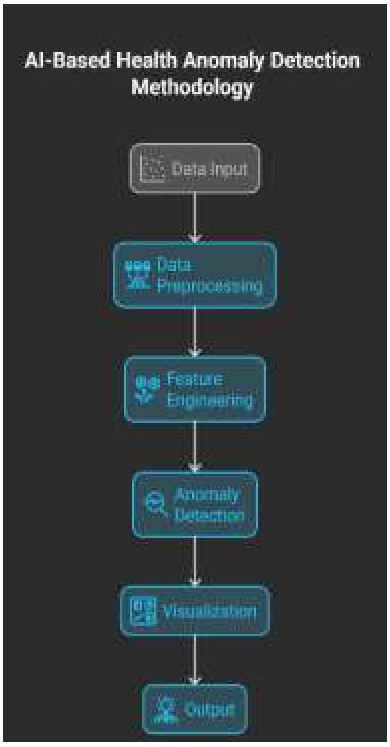

Fig: 1

V. ANOMALY DETECTION TECHNIQUES

To enhance the precision of detection, the framework utilizes a hybrid methodology that merges statistical rigor, unsupervised machine learning, and behavioral modeling. This multi-layered approach ensures that the system can identify both blatant outliers and more nuanced irregularities within wearable health data.

- **Statistical Baseline (Z-score)**: The system employs Z-score analysis to monitor individual parameters, such as heart rate and sleep duration, flagging any data points that exceed a defined threshold as extreme deviations.
- **Multivariate Machine Learning (Isolation Forest)**: An Isolation Forest model is implemented to detect anomalies within a multidimensional context. By isolating data points that exhibit distinct feature patterns, this unsupervised technique identifies complex irregularities that simple threshold-based methods often overlook.
- **Behavioral Analysis**: To account for individual physiological differences, the system incorporates behavioral modeling that compares current metrics against a user's specific historical baseline. This process uses rolling averages and variability measures to establish a dynamic, personalized window for anomaly detection.
- **Temporal Trend Analysis (Prophet)**: For deeper longitudinal insight, an optional component utilizes the Prophet forecasting model. This allows the system to model expected temporal trends and identify significant deviations between predicted and actual values.

By integrating these diverse analytical techniques, the system achieves a robust capability for uncovering both obvious and subtle health patterns in real-world fitness data

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 788-797

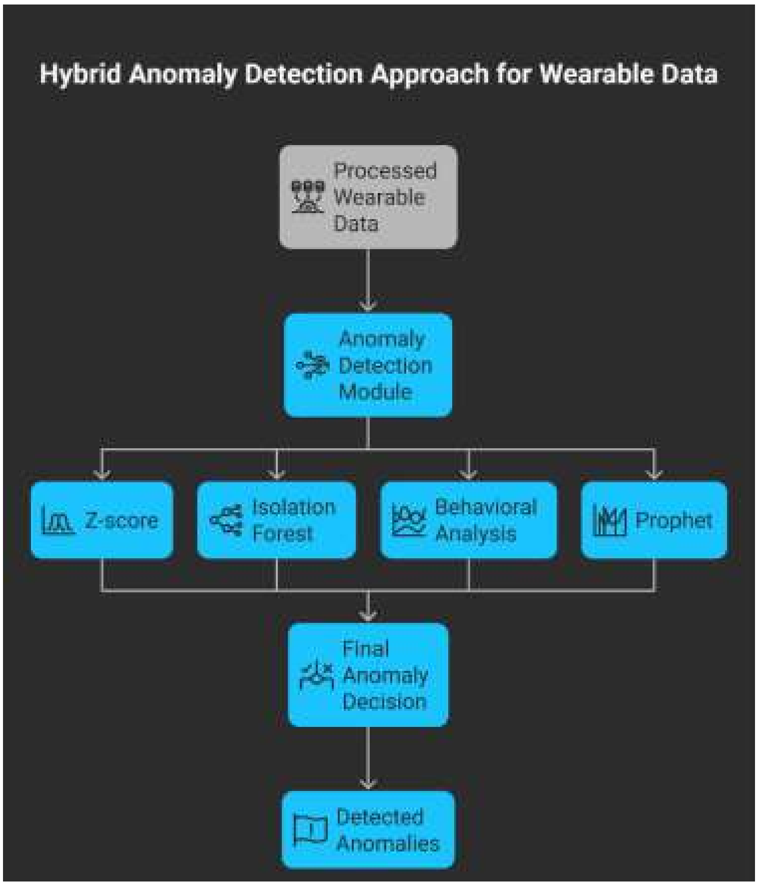

Fig 2

VI. SYSTEM ARCHITECTURE

The proposed system utilizes a modular architecture that decouples the user interface from the backend logic and analytical engines, a design choice that prioritizes both scalability and long-term maintainability. By separating these concerns, the framework ensures a streamlined data flow from raw input to final visualization.

- **Frontend Layer**: Developed using Streamlit, this layer serves as the primary touchpoint for users to upload datasets and interact with the results. It renders a structured, user-friendly interface where trends, health graphs, and anomaly indicators are displayed in real-time.
- **Backend Layer**: Built with FastAPI, this layer functions as the system's central controller, managing API communications and orchestrating the execution of various analytical tasks. It handles requests from the frontend and ensures they are efficiently routed to the appropriate processing modules.
- **Data Processing Layer**: This layer is divided into specialized modules dedicated to data cleaning and enrichment. The preprocessing module handles normalization and handles missing values, while the feature engineering module calculates derived metrics, such as rolling averages and variability measures, to provide deeper physiological context.

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 788-797

- **Anomaly Detection Module**: Serving as the analytical core, this module synthesizes several techniques, including Z-score, Isolation Forest, and behavioral analysis, to flag irregularities. It also includes an optional Prophet-based component for more complex time-series trend forecasting.

This interconnected design offers the flexibility needed for future expansions, such as the integration of live data streams or the addition of more sophisticated machine learning architectures.

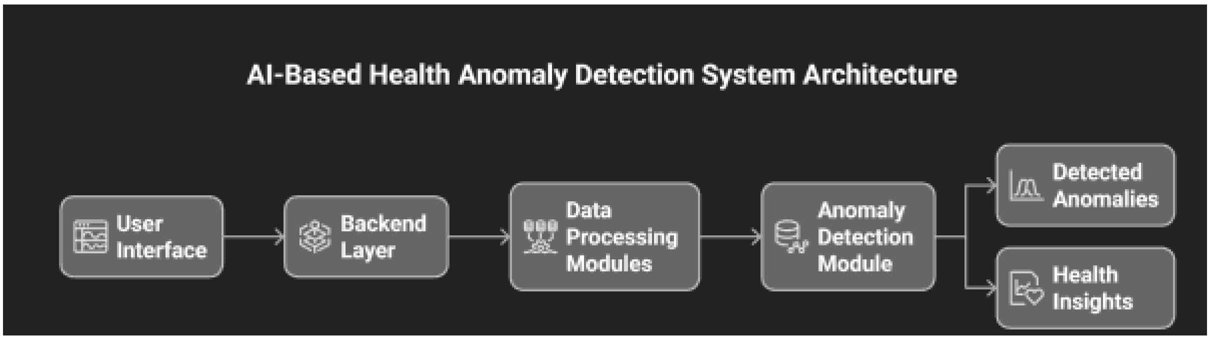

Fig 3

VII. IMPLEMENTATION

The technical execution of the proposed framework relies on a robust Python-based stack, chosen for its efficiency in data science, machine learning, and rapid application deployment.

- **Data Preprocessing**: We utilized the Pandas library to manage the initial data intake and cleaning. This stage involves reading CSV files and refining the dataset by resolving missing entries, standardizing column headers, and performing datetime conversions to facilitate time-series analysis.
- **Feature Engineering**: Using native Pandas operations, the system computes rolling statistics, including moving averages and standard deviations, over predefined temporal windows. These derived metrics are essential for capturing subtle fluctuations and short-term trends in a user's physiological behavior.
- **Anomaly Detection Core**: The detection logic combines statistical and machine learning methodologies. NumPy-based operations are used to calculate Z-scores for identifying extreme outliers in individual parameters. For more complex, multivariate irregularities, an Isolation Forest model is implemented via the Scikit-learn library. Furthermore, behavioral analysis relies on rolling averages to compare real-time metrics against historical trends, while Prophet is available as an optional module for advanced time-series modeling.
- **Backend Orchestration**: The backend is powered by FastAPI, which acts as the system's central nervous system. It manages the flow of information by receiving data requests, coordinating the execution of the detection modules, and relaying the final results back to the user interface.
- **Frontend and Visualization**: To provide an accessible user experience, the frontend is built with Streamlit. This interface allows for seamless file uploads and translates

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 788-797

complex analytical outputs into intuitive line charts and highlighted anomaly markers, ensuring the findings are easy for the end-user to interpret.

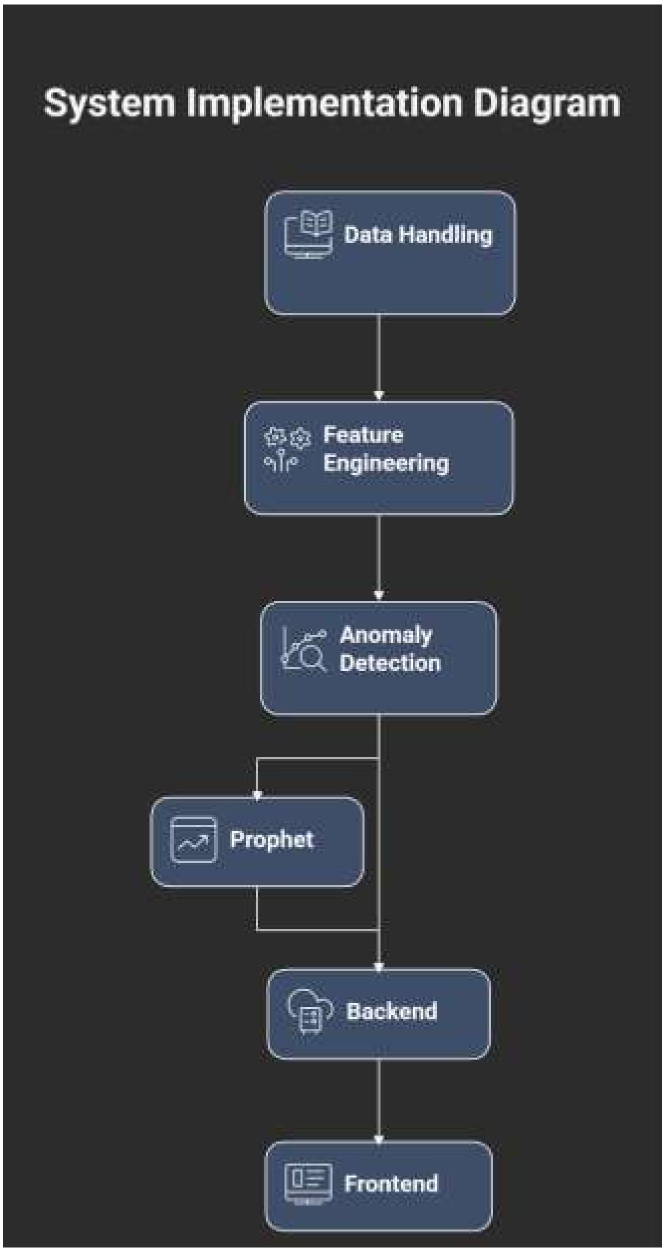

Fig 4

By integrating these diverse libraries and frameworks, the system maintains a high standard of data processing speed and detection accuracy within a user-centric environment.

VIII. RESULTS AND ANALYSIS

The proposed system underwent testing with wearable fitness datasets, focusing on core parameters such as heart rate, step count, and sleep duration. The framework successfully processed these inputs and generated valuable health insights through integrated anomaly detection and visualization.

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 788-797

- **Data Integrity**: The preprocessing module effectively cleaned and standardized the dataset, resolving missing entries and inconsistent column formats to ensure a reliable foundation for subsequent analysis.
- **Detection Performance**: The system's multi-layered analytical approach successfully pinpointed irregularities within the data. While the Z-score method isolated extreme individual outliers, the Isolation Forest model successfully identified more complex, multivariate anomalies. Furthermore, the inclusion of behavioral analysis added a layer of personalization by flagging deviations from a user's own recent trends.
- **Visual Interpretation**: Through the interactive dashboard, physiological trends were rendered into intuitive graphical formats, with detected anomalies clearly highlighted to facilitate easier identification of irregular patterns.
- **System Robustness**: By synthesizing various detection methodologies, the system demonstrated improved accuracy and resilience, successfully capturing both blatant outliers and subtle physiological shifts that might otherwise be overlooked.

Ultimately, these results suggest that the proposed framework offers a practical and dependable solution for the intelligent analysis of wearable fitness data.

IX. ADVANTAGES

- **Enhanced Precision**: By merging statistical methodologies with machine learning, the system significantly increases the overall accuracy of anomaly detection.
- **Multivariate Processing**: The architecture supports multivariate analysis, which enables the simultaneous identification of irregularities across several health parameters rather than evaluating them in isolation.
- **Personalized Insights**: Through behavioral pattern analysis, the system offers detection tailored to individual user trends, accounting for personal physiological baselines.
- **Intuitive Interpretation**: The platform delivers clear, interactive visualizations that simplify the interpretation of complex health data for the end-user.
- **Scalable Architecture**: The modular design ensures the system is highly scalable, facilitating future expansions such as the integration of real-time data streams or additional analytical models.

X. LIMITATIONS

- **Offline Data Dependency**: The current framework relies exclusively on static dataset inputs and lacks direct, real-time integration with wearable device APIs.
- **Sensitivity to Data Quality**: The precision of anomaly detection is heavily dictated by the cleanliness and completeness of the provided input data.
- **Computational Constraints**: The Prophet-based trend analysis remains an optional feature due to its high computational demands, which may limit its performance on hardware with lower processing power.

ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 788-797

- **Foundational Personalization**: While the system utilizes behavioral baselines, it does not yet incorporate more sophisticated personalization techniques beyond these fundamental pattern evaluations.

XI. FUTURE SCOPE

- **Real-time Synchronization**: Future iterations aim to move beyond static datasets by integrating directly with wearable device APIs, allowing for live data streaming and continuous health monitoring.
- **Cloud Integration**: Transitioning the system to a cloud-based infrastructure would facilitate greater scalability, centralized data storage, and seamless remote access for both users and healthcare providers.
- **Advanced Model Implementation**: There is significant potential to incorporate more sophisticated deep learning architectures—such as LSTMs or Autoencoders—to further refine detection accuracy in complex time-series data.
- **Prescriptive Analytics**: Beyond mere detection, the system could evolve to include personalized recommendation engines that suggest specific lifestyle adjustments based on identified health patterns.
- **UI/UX Optimization**: Continued refinement of the user interface will focus on enhancing interactivity and usability, ensuring that complex analytical insights remain accessible and engaging for a non-technical audience.

XII. CONCLUSION

This study has detailed an effective framework for interpreting wearable fitness data through a hybrid methodology that synthesizes statistical rigor, machine learning, and behavioral pattern evaluation. By integrating diverse analytical layers, including Z-score analysis for extreme outliers and Isolation Forest for complex multivariate patterns, the system successfully identifies both blatant and subtle physiological deviations. Robust preprocessing and feature engineering stages ensure the integrity of the data, while the interactive dashboard translates these technical findings into interpretable health intelligence for the user. The findings confirm that a multi-layered detection strategy significantly enhances accuracy and reliability over traditional, single-method systems. Ultimately, this architecture offers a practical and scalable solution for converting raw sensor output into actionable insights, providing a valuable foundation for the future of personal health monitoring and proactive preventive care.

REFERENCES

[1] A. Reiss and D. Stricker, "Introducing a New Benchmarked Dataset for Activity Monitoring," *IEEE International Symposium on Wearable Computers*, 2012.
[2] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep Learning for Sensor-based Activity Recognition: A Survey," *Pattern Recognition Letters*, vol. 119, pp. 3–11, 2019.
[3] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, pp. 413–422, 2008.
[4] R. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, OTexts, 2018.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-5 | May-2026 | Page 788-797

[5] S. J. Taylor and B. Letham, "Forecasting at Scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.
[6] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Computing Surveys*, vol. 41, no. 3, 2009.
[7] H. Gjoreski, M. Luštrek, and M. Gams, "Accelerometer Placement for Posture Recognition and Fall Detection," *Proceedings of the International Conference on Intelligent Environments*, 2011.
[8] Fitbit Fitness Tracker Dataset, Kaggle.
[9] Fitbit Fitness Tracker Data Analysis Notebook, Kaggle.
[10] Fitbit Fitness Tracker Data Exploration, Kaggle.
[11] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, 2009.
[12] Scikit-learn Documentation.
[13] Pandas Documentation.
[14] FastAPI Documentation.
[15] Streamlit Documentation.