

**National University of Computer & Emerging Sciences
Karachi Campus**



Machine Learning
Date of submission: 12 December 2023

**Advancements in Elderly Care: A Machine Learning
Approach to Automated Fall Detection Using Smartphone
Sensors**

21K-4703 ALI RAZA
21K-3100 MUHAMMAD SAMEED
21K-4736 SYED SAADULLAH HUSSAINI

(BAI-5A)

Supervisor: Dr. Muhammad Farrukh Shahid

Abstract:

The rise of computing and sensor technology has paved the way for innovative healthcare applications, particularly in addressing the challenges of elderly care and fall detection. Falls among the elderly pose significant risks, necessitating automated systems for early detection and prevention. This report outlines the development of such a system, utilizing smartphone and smartwatch inertial sensors, and introduces a machine learning framework for identifying and classifying elderly activities, with a specific focus on fall detection.

The methodology involves Evaluation metrics such as accuracy, precision, recall, and F1-score are employed to assess model performance. The report delves into the impact of feature selection on generalization and effectiveness in distinguishing falls from non-fall activities. Furthermore, it provides visual insights into accelerometer data distributions, shedding light on movement patterns relevant to fall detection. Beyond the technical aspects, the report narrates a compelling story of how machine learning applications can significantly impact elderly well-being, highlighting the potential to save lives and enhance the quality of life for an aging population.

Introduction

As the elderly population globally rises, addressing age-related health concerns, particularly the risk of falls, becomes imperative. Falls, common among the elderly, demand real-time monitoring solutions. Integrating smartphone and smartwatch inertial sensors offers a promising avenue for effective, cost-efficient fall detection and prevention. The report's significance lies in outlining the development and evaluation of an automated fall detection system using machine learning techniques, leveraging widely accessible sensors. Two feature engineering techniques will be explored, employing statistical feature extraction and direct utilization of raw data. Machine learning models, including Random Forest, Logistic Regression, and Support Vector Machines, will be evaluated for accuracy in classifying activities and detecting falls. Visualizations of sensor data distributions will provide analytical insights. The report structures a comprehensive exploration of the fall detection system, covering literature review, methodology, results interpretation, and conclusions, with a focus on technical aspects, challenges, solutions, and outcomes.

Methodology

In formulating our methodology, we harness a meticulously curated multimodal dataset dedicated to fall detection. This dataset seamlessly integrates data from smartphone inertial sensors, complemented by infrared signals and trials, bolstering its comprehensiveness. Our method places considerable emphasis on the foundational steps of data preprocessing and feature engineering, pivotal for effective machine learning model training and assessment.

Data Utilization

In this study, we opted not to conduct primary data collection and instead utilized the UP-Fall Detection Dataset, a well-established resource in fall detection research.^[1] This dataset, purposefully curated for algorithm development, encompasses a variety of activities and falls captured through different sensors, facilitating robust applications of machine learning. Including readings from

accelerometers and infrared sensors, it provides a comprehensive perspective essential for accurate fall detection. Acquired ethically from a reputable source, the dataset's thoroughness allows for in-depth exploration of machine learning techniques. Properly acknowledged, the UP-Fall Detection Dataset serves as the primary data source for analyses, feature engineering, model training, and validation. This methodology ensures elevated data quality and relevance for fall detection while circumventing the necessity for primary data collection.^[1]

Data Preprocessing

The initial phase of data preprocessing involves the normalization and standardization of data from an array of sensors, ensuring consistency and comparability. Specifically, histograms of accelerometer readings from various body locations uncover distinct peaks and patterns, guiding our feature engineering approaches.

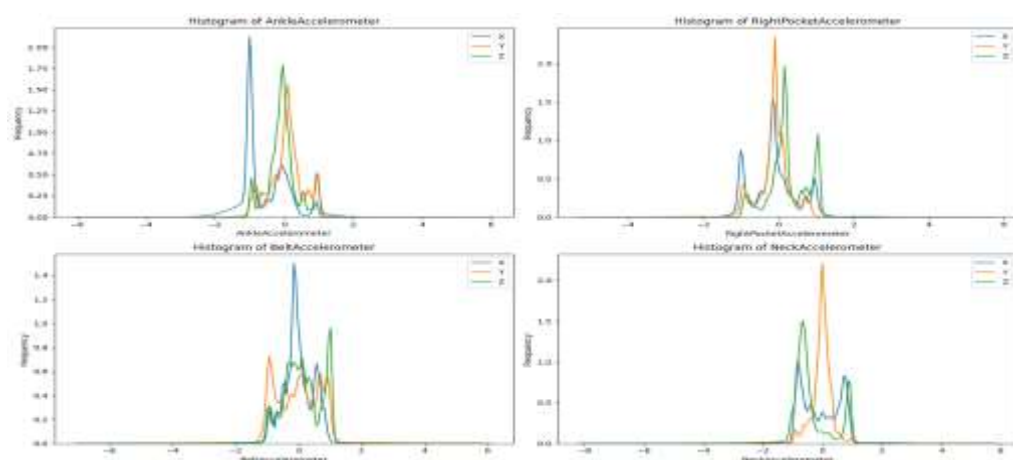
Feature Engineering

The feature engineering approach involves a comprehensive feature set, incorporating all features except the 'activity' column. These are treated as predictors (X), with activity labels serving as the target variable (Y). This extensive feature set is crafted to capture subtle differences in sensor readings across a spectrum of activities, enriching the depth of our fall detection model.

Visual Data Analysis

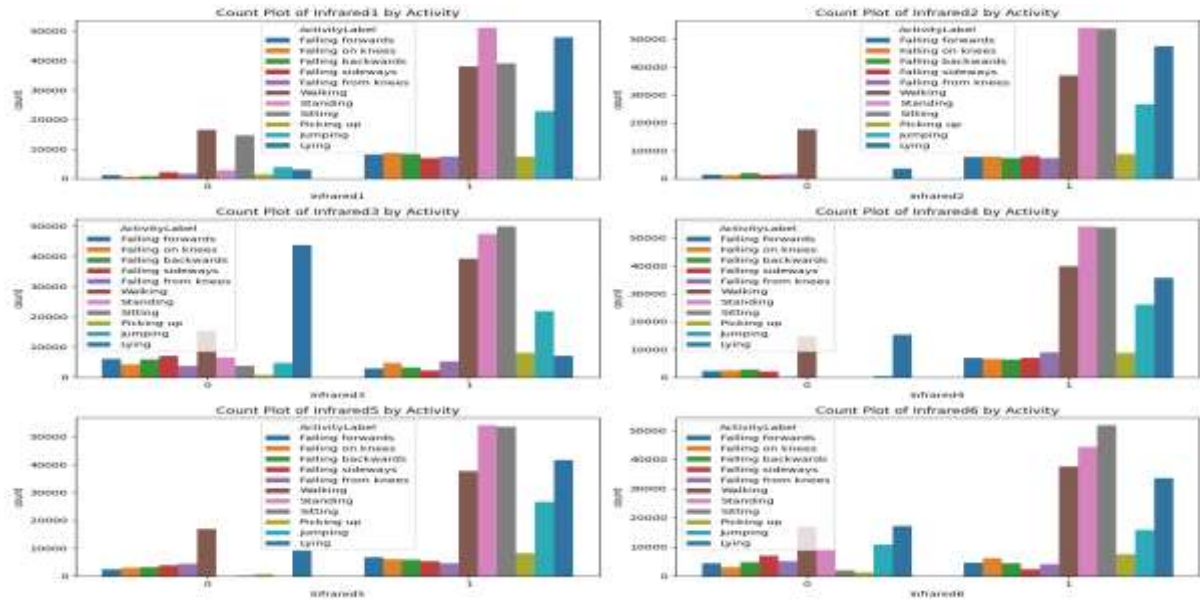
The visual analysis begins with histograms representing the distribution of accelerometer data across different sensors attached to the body, including the ankle, belt, right pocket, and neck. Each histogram shows three lines, each representing a different axis of acceleration (X, Y, Z). These graphs provide a visual representation of the frequency and intensity of movements captured by the sensors.

- **Ankle Accelerometer:** Sharp peaks suggest rapid movement changes, likely capturing the impact of footfalls or sudden shifts in position.
- **Belt Accelerometer:** The distribution is less sharp than the ankle, indicating more varied motion patterns, possibly related to the torso's movement during various activities.
- **Right Pocket Accelerometer:** This graph likely captures a mix of leg and upper body movement, showing multiple prominent peaks.
- **Neck Accelerometer:** The data shows less variability, which may indicate that the neck experiences more uniform motion during the activities monitored.



[Graph of Ankle Accelerometer, Belt Accelerometer, Right Pocket Accelerometer, and Neck Accelerometer]

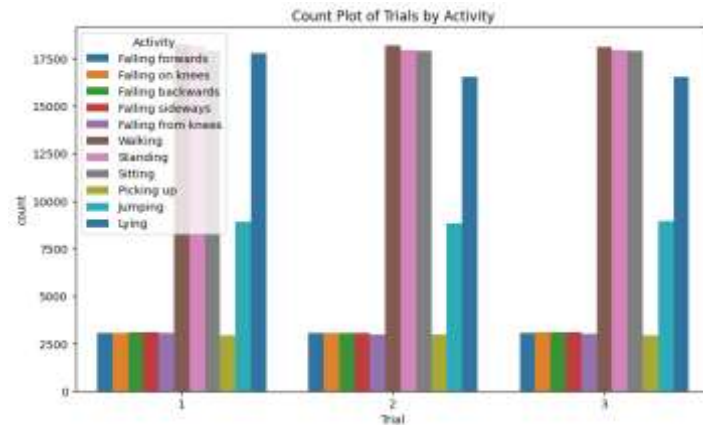
Our methodology includes a meticulous analysis of visual data. Count plots corresponding to infrared signals, cross-referenced with activity labels, offer a visual representation of activity frequency against binary states of infrared sensors. This analysis sheds light on associations between specific activities (e.g., 'Falling forwards,' 'Falling sideways,' and 'Lying') and infrared sensor states, providing valuable insights for the fall detection mechanism.



[Count plots of Infrared Sensors observations by Activity]

- **Infrared1 and Infrared2:** The depicted plots highlight a notable occurrence of "walking" and "lying" activities when the sensor state is '1'. This suggests effective positioning of these sensors for detecting an individual in an upright or lying position. A distinct difference in "lying" occurrences between the two states may signify the sensor's sensitivity to postures associated with rest or potential falls.
- **Infrared3 and Infrared4:** These sensors show a considerable count of "walking" and "standing" activities in state '1'. Additionally, "Falling forwards" and "Falling sideways" are more prevalent in this state, suggesting proficiency in detecting falls, especially in the sensor's oriented direction. The significant contrast in counts for these falling activities between states highlights the sensors' effectiveness in fall detection.
- **Infrared5 and Infrared6:** Both states exhibit a notable presence of "walking," "standing," and "lying" activities, with a substantial number of "lying" instances in state '1'. This could indicate these sensors' ability to detect a horizontal orientation. Furthermore, the significant occurrence of "falling from knees" in state '1' for Infrared5 hints at its role in recognizing falls from a kneeling position.

Additionally, count plots illustrating trials by activity reveal the distribution of activities across different trials. This analysis not only underscores the consistency or variance in collected data for each activity but is also instrumental in gauging the reliability of the dataset for robust machine learning model training.



[Graph: Count Plot of Trials by Activity]

- **Consistent Trial Distribution:** Similar patterns in activity distribution across trials suggest uniform data collection methods and a balanced representation, critical for training reliable machine learning models.
- **Prominence of "Walking" and "Lying":** "Walking" and "lying" activities emerge as frequently occurring across all trials, potentially indicating a focus or ease of capture during data collection.
- **Balanced Falling Activities:** Various falling activities ("falling forwards," "falling on knees," "falling backwards," "falling sideways," "falling from knees") exhibit a balanced representation, facilitating the development of a model capable of recognizing diverse fall types.
- **Infrequent Activities:** Activities like "picking up" and "jumping" show lower occurrences, suggesting either a lesser focus during data collection or intrinsic infrequency.
- **Stability in "Standing" and "Sitting":** Consistent presence of "standing" and "sitting" activities across trials implies their stable and recurrent nature within the dataset.

This yields valuable insights into the dataset's structure, revealing activity distribution nuances across trials. Such insights aid in detecting imbalances or biases that may impact model training and performance. Moreover, it verifies data robustness, ensuring a model trained on this dataset generalizes effectively to real-world scenarios.

Model Training and Evaluation

We opt for a diverse set of machine learning and deep learning models, encompassing decision trees, support vector machines, random forests, and neural networks. To ensure the generalizability of our models, we apply cross-validation techniques across the dataset.

- For the Random Forest Classifier, its performance may have either benefited from the information richness in the raw data or suffered if the raw data contained noise.
- In the case of Logistic Regression, challenges might have arisen in navigating the high-dimensional space if regularization was not appropriately applied.
- The effectiveness of SVM could hinge on the choice of kernel and its capacity to manage the high-dimensional feature space proficiently.

Performance evaluation employs metrics such as accuracy, precision, recall, and the classification report. Notably, we conduct a comparative analysis of model accuracy after applying the feature engineering technique. This nuanced examination informs us about the impact of feature selection on model performance and aids in identifying the most effective features and models for fall detection.

Results:

This section highlights outcomes from assessing diverse machine learning models for human fall detection using smartphone inertial sensors with the UP-Fall Detection Dataset.^[1] Key metrics include accuracy, precision, recall, and F1-score.

Random Forest Classifier

Optimized with a hyperparameter grid search, the Random Forest model achieved near-perfect accuracy in cross-validation and approximately 99.98% on the test set. Outstanding F1-scores (1.00) across all classes underscore its exceptional classification proficiency.

Logistic Regression:

Optimized for the regularization strength parameter (C), the Logistic Regression model, with a C value of 10, delivered high cross-validation accuracy (95.91%) and a test set accuracy of 96.27%. Notably, precision and recall contributed to an overall F1-score ranging from 0.85 to 1.00.

Support Vector Machine (SVM)

The SVM, optimized through a grid search, achieved impressive cross-validation accuracy (99.58%) and maintained excellence on the test set (99.87%). Consistent high precision, recall, and F1-scores across classes showcased robust classification capabilities.

Deep Learning Model

A neural network, trained over 20 epochs, peaked at 93.47% validation accuracy. With low test loss (0.2152) and a test accuracy of 93.87%, the model demonstrated effectiveness in generalizing to new data.

Model Comparison

Comparatively, Random Forest and SVM models stood out for superior accuracy and robustness, while Logistic Regression and neural network models performed well. Overall high accuracy rates underscore machine learning's potential in enhancing fall detection for the safety and independence of the elderly. Further optimization could enhance the potential of Logistic Regression and neural network models.

Conclusion:

This technical report illustrates the effectiveness of machine learning methods in creating a fall detection system with smartphone sensors. In-depth analysis reveals the Random Forest and SVM models showcase superior accuracy, showcasing their viability for practical healthcare use. The outcomes, supported by strong precision and recall, affirm the resilience of the employed feature engineering and model optimization. These results open avenues for enhanced independent health monitoring for the elderly, ensuring a harmonious blend of safety and autonomy. Consequently, this study establishes a sturdy groundwork for future progress in fall detection technologies.

References

1. Lourdes Martínez-Villaseñor, Hiram Ponce, Jorge Brieva, Ernesto Moya-Albor, José Núñez-Martínez, Carlos Peñafort-Asturiano, “UP-Fall Detection Dataset: A Multimodal Approach”, *Sensors* 19(9), 1988: 2019, doi:10.3390/s19091988.