# Enhancing Minority Class Prediction in Wearable Sensor-Based Activity Recognition Using SMOTE Oversampling

Sarmini[1,*], Chyntia Raras Ajeng Widiawati[2], Ika Romadoni Yunita[3]

[1,3]Infomation System, Computer Science Faculty, Universitas Amikom Purwokerto, Indonesia

[2]Information Technology, Computer Science Faculty, Universitas Amikom Purwokerto, Indonesia

(Received: November 1, 2024; Revised: December 15, 2024; Accepted: February 10, 2025; Available online: April 1, 2025)

**Abstract**

Wearable sensor-based activity recognition has become increasingly important in various domains, particularly healthcare and sports. However, a significant challenge in this field is the issue of class imbalance, where minority activity classes are underrepresented compared to majority classes in datasets. This imbalance leads to biased classifiers that struggle to accurately identify rare but critical activities, which is especially problematic in health monitoring scenarios. This study evaluates the effectiveness of the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance in the mHealth dataset, which contains multi-sensor data from wearable devices placed on the chest, left ankle, and right lower arm. We employ the XGBoost classifier combined with SMOTE oversampling to improve recognition performance for minority classes. Model evaluation is conducted using precision, recall, F1-score, Area Under the Precision-Recall Curve (AUC-PR), ROC curve, and calibration analysis. The results demonstrate that applying SMOTE improves minority class recall from 0.75 to 0.85 and F1-score from 0.796 to 0.865, despite a slight decrease in overall accuracy from 97% to 96.5%. The AUC-PR also increases from 0.81 to 0.88, indicating a better balance in detecting minority and majority classes. Calibration curves reveal that probability estimates still require refinement to be more reliable for decision-making. This study confirms the efficacy of SMOTE in mitigating class imbalance in wearable sensor-based activity recognition and provides valuable insights for developing more accurate and fair health monitoring systems.

*Keywords*: Wearable Sensors, Activity Recognition, Class Imbalance, SMOTE, XGBoost

## 1. Introduction

Wearable sensor-based activity recognition has gained significant importance across various applications, especially in healthcare and sports. In healthcare, wearable sensors are transforming the way health monitoring is carried out by providing continuous real-time data on vital signs. These sensors enable early detection of potential health complications, thus improving healthcare efficiency [1]. By tracking essential indicators such as heart rate, blood pressure, and body motion, wearable sensors contribute significantly to personal health management and proactive interventions [1]. For instance, wearable devices have become a cornerstone in chronic disease management, allowing healthcare professionals to track patient health outside of clinical settings, thus reducing hospital visits and enabling timely medical responses.

In addition to healthcare, wearable sensors have also made a transformative impact on sports performance analysis. These devices provide objective metrics on athletes' physiological parameters, including muscle activity and movement patterns, which are essential for optimizing training strategies [2]. Wearable technology in sports not only helps track performance but also provides insight into recovery, endurance, and injury prevention. This capability is especially critical for athletes looking to fine-tune their performance and minimize the risk of injury. Furthermore, the development of advanced materials and sensor technologies has enhanced the reliability and usability of these devices, allowing them to be used in a variety of environmental conditions without disrupting the user's activities [3]. The integration of smart materials and artificial intelligence has further amplified the capabilities of wearable sensors, enabling more nuanced health insights and performance tracking [4]. Despite the promising advancements in wearable

*Corresponding author: Sarmini (sarmini@amikompurwokerto.ac.id)
DOI: https://doi.org/10.47738/ijaim.v5i1.95

ISSN 2776-8007


sensor technology, challenges persist in the domain of activity recognition. One of the most significant obstacles is the problem of class imbalance, particularly in activity classification tasks where some activities are less frequently represented in the dataset. Class imbalance is a pervasive issue in machine learning, as classifiers tend to favor the majority class, often leading to suboptimal performance in identifying minority activities [5]. This issue is compounded by noise interference, which further complicates the accurate identification of minority class instances, making the model biased toward the majority class [5].

Class imbalance not only affects the accuracy of predictions but also undermines the model's ability to detect rare but critical activities. This can have profound implications, particularly in applications like healthcare monitoring, where detecting minor but significant activities such as sudden changes in heart rate or unexpected movements is crucial. Consequently, addressing class imbalance has become a critical area of focus in activity recognition research.

Several strategies have been developed to address this imbalance, with oversampling and under sampling being the most widely employed techniques. These methods aim to adjust the dataset so that the minority class is adequately represented, allowing classifiers to learn from a more balanced set of data [6]. Among these methods, the Synthetic Minority Over-sampling Technique (SMOTE) has gained considerable attention for its ability to generate synthetic samples for the minority class, thus mitigating the bias toward the majority class [7]. Furthermore, hybrid approaches that combine oversampling with feature selection and advanced algorithms have shown promising results in improving classification accuracy for imbalanced datasets. These techniques, including class-aware loss functions, are designed to assign greater weight to the minority class, further enhancing its recognition within the learning process [8].

However, while the use of SMOTE and other oversampling techniques has shown efficacy in numerous applications, their use in wearable sensor-based activity recognition remains underexplored. Although studies like those by Sánchez-Gutiérrez and González-Pérez [9] have demonstrated the effectiveness of SMOTE in imbalanced tabular datasets, the application of these methods specifically to wearable sensor data is still lacking [9]. Their work primarily focuses on Generative Adversarial Networks (GANs) for synthetic data generation, which may not directly translate to the particular challenges presented by wearable sensor data. This highlights a significant gap in the literature, particularly in the context of wearable sensors, where the interplay between class imbalance, data noise, and sensor-based features requires further investigation.

Moreover, research has also demonstrated that standard classification metrics such as accuracy can be misleading in imbalanced datasets, as they fail to provide a complete picture of model performance. Accuracy tends to be high in models biased toward the majority class but may not reflect the model's true performance when it comes to identifying minority class instances. This issue is particularly prominent in applications such as healthcare diagnostics or intrusion detection systems, where detecting rare events is far more important than simply achieving high overall accuracy [10]. In these scenarios, precision and recall become more informative metrics, providing insight into the model's ability to detect positive instances and avoid false positives [11].

Precision, recall, and F1-score are commonly used to evaluate classifier performance in imbalanced datasets. Precision measures the proportion of true positive predictions among all positive predictions, while recall assesses how well the model identifies all relevant instances [12], [13]. The F1-score, which balances precision and recall, offers a more holistic view of performance, especially in critical applications where both false positives and false negatives can have serious consequences [14]. Furthermore, calibration is essential for improving decision-making in systems that rely on probability estimates, such as medical predictions, where inaccurate probabilities can lead to misleading conclusions [15].

The current literature has made significant progress in addressing these challenges, but there is still a need for more focused studies on the application of SMOTE and other oversampling techniques within the context of wearable sensor data. For instance, while research by Li et al. [16] has underscored the importance of robust evaluation metrics, such as precision, recall, and F1-score, their findings are not specific to wearable sensor data [16]. Furthermore, studies by Kim and Hwang [17] have demonstrated that SMOTE can effectively handle class imbalance in various datasets, but there remains a lack of comparative studies focused specifically on wearable sensor applications [17]. As wearable sensor technology continues to evolve and find new applications in healthcare, fitness monitoring, and beyond, addressing class imbalance and improving the reliability of predictive models will be crucial for their success.

 ISSN 2776-8007


Wearable sensor-based activity recognition systems hold great promise for applications in both healthcare and sports, providing continuous and detailed insights into a person's health and performance. However, overcoming challenges such as class imbalance is essential for ensuring that these systems can accurately detect rare but critical activities. Techniques such as SMOTE offer a promising solution to this problem, and further research into their application in wearable sensor data can help improve classification performance and make these systems more reliable and effective in real-world scenarios. As the field continues to evolve, integrating advanced data balancing methods with robust model evaluation strategies will be key to unlocking the full potential of wearable sensor technology for health and performance monitoring.

**2. Literature Review**

## 2.1. Advances in Wearable Sensor-Based Activity Recognition and Integration

Wearable sensor-based activity recognition has gained substantial attention due to its broad applications in healthcare, sports, and other fields. In particular, deep learning models have emerged as a key trend in improving activity recognition systems. These models have shown considerable promise in identifying complex patterns in sensor data. For instance, Sa-Nguannarm et al. [18] utilized Bidirectional Long Short-Term Memory (Bi-LSTM) networks to analyze stress behavior, showcasing the effectiveness of Recurrent Neural Networks (RNNs) for handling sequential data. Similarly, Yadav et al. [19] introduced ARFDNet, a system that integrates activity recognition with fall detection, achieving high accuracy and specificity. This highlights the growing trend of integrating multiple functionalities into wearable sensor systems, particularly for applications in elder care.

Another key development in wearable sensor-based activity recognition is the integration of wearable sensors with advanced data collection and processing techniques. Research by Guerra et al. [20] demonstrated the use of RGB-D cameras in Ambient Assisted Living (AAL) environments, where skeletal representations were used to enhance activity recognition while respecting user privacy. This approach facilitates the development of more user-friendly solutions for healthcare applications. Additionally, Nouriani et al. [21] explored the use of low-cost wearable devices for real-time mobility monitoring in patients with movement disorders. This research highlights the potential for wearable sensors to expand beyond clinical settings, making health monitoring more accessible and practical in everyday life, particularly for patients with chronic conditions.

In addition to deep learning and sensor integration, various novel approaches have been explored to improve Human Activity Recognition (HAR). Asl et al. [22] provided a statistical overview of human activity categorization using wearable sensor data, establishing a foundational understanding of the subsets of activities that need to be recognized. Brard et al. [23] proposed a model that utilizes hip rotation data to accurately identify walking phases and distinguish them from non-walking activities, demonstrating an innovative approach to sensor utilization in real-world settings. These studies, along with others that integrate artificial intelligence and multimodal sensing, reflect the shift toward optimizing HAR by utilizing diverse sensor types and advanced AI techniques. AI has shown significant potential in improving model performance by incorporating complex physiological data. For example, Ramirez et al. [24] used skeleton features and data augmentation to enhance activity recognition accuracy, while Vijayvargiya et al. [25] employed surface electromyography (sEMG) signals to classify lower limb activities, showcasing AI's role in refining wearable sensor systems.

## 2.2. Class Imbalance in Activity Recognition

Despite the advancements in wearable sensor technology, a persistent challenge in activity recognition remains: class imbalance. Class imbalance refers to the uneven distribution of instances across different classes, which significantly affects the performance of machine learning models. Zhang et al. [26] discuss the recognition accuracy issues caused by class imbalance, particularly in activity recognition tasks where some activities may be infrequent, leading to insufficient training on minority classes. This imbalance can lead to models that are biased toward the majority class, undermining their ability to accurately classify rare but critical activities.

Lv and Liu [27] emphasize the impact of class overlap, where different classes intersect in feature space, exacerbating the class imbalance issue. They argue that class overlap results in the misclassification of minority class instances, as models tend to favor the majority class. This issue is more complex than simple class imbalance and requires innovative solutions to ensure that minority classes are adequately represented and recognized.

Bako et al. [28] highlight the performance degradation in minority class recognition, noting that standard classifiers struggle to predict minority classes accurately when class imbalance is present. They advocate for adaptive synthetic sampling techniques like ADASYN, which generate synthetic samples to address class imbalance and improve model performance, particularly for underrepresented classes.

## 2.3. Addressing Class Imbalance: Techniques and Strategies

Several studies explore techniques to address class imbalance in HAR. Hosseini et al. [29] argue that shallow learning approaches are insufficient in dealing with class imbalance, particularly in deep learning frameworks. They suggest that more advanced techniques need to be developed to handle class imbalance effectively in deep learning models. Alharbi et al. [6] provide an overview of various sampling strategies that have been employed to tackle imbalanced data in HAR, emphasizing the importance of tailored techniques to improve model performance. Their research underscores the need for careful handling of imbalanced data to ensure accurate recognition of rare activities.

Kumar [30] discusses the generalizability of insights regarding class imbalance in machine learning models, noting that high overall accuracy can mask poor performance for minority classes. This finding highlights a critical flaw in many machine learning models that assume balanced class distributions. Addressing these imbalances actively is necessary to improve classification accuracy, particularly in applications where minority class detection is crucial.

## 2.4. SMOTE in Activity Recognition

One of the most widely studied approaches for addressing class imbalance is the SMOTE, which generates synthetic samples for minority classes to improve classifier sensitivity. Several studies have explored the application of SMOTE in HAR, particularly in the context of wearable sensors. Elreedy et al. [31] provided a theoretical framework for understanding SMOTE, detailing how synthetic samples are generated by interpolating between minority class instances and their nearest neighbors. This theoretical approach ensures a more informed handling of class imbalance, providing a foundation for its successful application in wearable sensor data.

Other studies have introduced hybrid approaches that combine SMOTE with other techniques. For instance, Hartono and Ongko [32] proposed a hybrid model that integrates SMOTE with ensemble methods to manage multi-class imbalanced datasets, showing significant improvements in predictive performance. Husain et al. [33] compared SMOTE with SMOTEENN (SMOTE combined with Edited Nearest Neighbors) and found that SMOTEENN outperformed standard SMOTE in maintaining accuracy and reducing error rates, particularly in complex regression models.

## 2.5. Challenges and Future Directions

While SMOTE and its variants have shown effectiveness in improving the performance of activity recognition models, challenges remain. One significant issue is class overlap, where synthetic samples may introduce additional overlap between classes, complicating classification tasks. Zhou et al. [34] demonstrated the versatility of SMOTE in applications such as motion-capturing systems, where combining oversampling methods with other strategies can improve activity recognition accuracy.

Addressing class imbalance in wearable sensor-based activity recognition is crucial for improving model performance, particularly in real-world applications where rare activities must be accurately identified. SMOTE and its variations offer promising solutions, but further research is needed to refine these techniques and overcome challenges like class overlap. As the field progresses, a combination of advanced oversampling methods, innovative sensor fusion techniques, and comprehensive model evaluation frameworks will be key to advancing the effectiveness of wearable sensor systems for activity recognition.

## 3. Methodology

Figure 1 illustrates the overall workflow employed in this study, providing a structured overview of the processes undertaken to develop and evaluate the classification model. This diagram helps to visualize the methodological framework and ensure a clear understanding of the research approach.

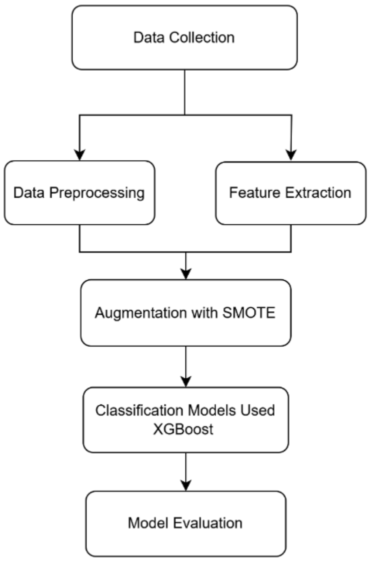


**Figure 1.** Research Methodology

### 3.1. Data Collection

The mHealth dataset is a comprehensive collection of physiological and motion data obtained from wearable sensors and made publicly available on Kaggle. It was recorded from human subjects outfitted with sensor devices positioned at three key body locations: the chest, left ankle, and right lower arm. Each device is equipped with a 3-axis accelerometer, gyroscope, and magnetometer, while the chest sensor also includes two ECG leads. These sensors continuously capture detailed measurements of the subjects' movements and physiological states as they perform various physical activities.

The dataset comprises 13 activity classes labeled from 0 to 12. Class 0 corresponds to periods of no activity or undefined behavior, while classes 1 through 12 represent specific physical actions such as standing, sitting, lying down, walking, stair climbing, waist bending, arm elevation, crouching, cycling, jogging, running, and jumping. Each data point is timestamped and annotated with an activity label, enabling supervised learning for activity recognition. However, the data distribution is imbalanced, with class 0 having a disproportionately high number of samples compared to other classes. The mHealth dataset is widely utilized in machine learning and healthcare research to build models that can recognize and monitor human activities, with applications in areas like fitness tracking, elderly care, and physical rehabilitation.

### 3.2. Data Preprocessing, and Feature Extraction Steps

Data preprocessing is a vital step in preparing the mHealth dataset for machine learning and analysis. The raw sensor data, collected continuously from various body-mounted devices, first undergoes cleansing to ensure quality and consistency. This involves checking for and handling missing or corrupt values, which may arise due to temporary sensor disconnections or recording errors. Each sensor signal such as acceleration, gyroscopic movement, and ECG is then standardized or normalized to bring the values to a comparable scale, preventing any single sensor from disproportionately influencing the model. Additionally, the continuous time-series data is segmented into smaller, fixed-length windows (e.g., 2 to 5 seconds) using a sliding window approach. Each segment is treated as a single data instance and is typically labeled based on the dominant activity within the window. This segmentation process is essential for converting raw streams into structured inputs suitable for supervised learning models.

 ISSN 2776-8007


Once the data has been preprocessed and segmented, the next critical step is feature extraction. This involves converting raw sensor readings into meaningful numerical summaries that capture the essence of the activity being performed. For each windowed segment, a range of time-domain features are computed, such as mean, standard deviation, min, max, signal magnitude area (SMA), and root mean square (RMS), which reflect the intensity and variability of movement. Additionally, frequency-domain features may be extracted using techniques like the Fast Fourier Transform (FFT), allowing analysis of signal energy distribution across frequency bands a useful representation for dynamic activities like jogging or jumping. In cases involving ECG data, physiological features like estimated heart rate and R-R intervals can also be included. Furthermore, sensor fusion techniques can be applied to combine signals from multiple devices (e.g., computing the magnitude of 3-axis accelerometer data), thereby enriching the feature set. The resulting feature vectors provide a robust and compact representation of the underlying physical activity, ready for input into classification models.

## 3.3. Explanation of the Applied SMOTE Oversampling Technique

SMOTE is a widely used and effective method for addressing class imbalance in machine learning datasets, particularly in classification tasks such as human activity recognition using the mHealth dataset. In this dataset, the distribution of activity classes is highly skewed, with class 0 (representing no activity) significantly outnumbering other classes like jumping, cycling, or arm movements. Such imbalances can lead to classifiers that are biased toward majority classes, reducing performance on the minority ones.

SMOTE addresses this issue by generating synthetic examples for the minority class rather than simply duplicating existing instances, which can lead to overfitting [35]. It achieves this by selecting a random data point from the minority class and identifying its k nearest neighbors (typically k = 5) within the same class. New synthetic samples are then created by interpolating along the linear paths between the original sample and its neighbors [36]. This technique enhances diversity within the minority class and encourages the learning of more generalizable decision boundaries.

SMOTE has been shown to improve classifier performance in various domains, especially in biological and medical contexts [37]. When applied to the mHealth dataset, SMOTE helps ensure that underrepresented activities like jumping or frontal arm elevation are more adequately modeled, leading to improved classification accuracy and greater fairness across activity types. However, while SMOTE mitigates bias and increases model robustness, it may also introduce noise or generate borderline samples that do not well represent the true minority class distribution [38]. To counter such limitations, hybrid methods combining SMOTE with under-sampling or other refinement techniques are often explored for more reliable and balanced outcomes [39].

## 3.4. Classification models used XGBoost

Extreme Gradient Boosting (XGBoost) algorithm was employed as the classification model to recognize physical activities from the mHealth wearable sensor dataset. XGBoost is a powerful ensemble learning technique based on gradient boosting decision trees, known for its accuracy, speed, and efficiency, especially with structured data and imbalanced datasets. It constructs an ensemble of weak learners (typically decision trees) in a stage-wise manner, where each new tree attempts to correct the prediction errors made by the preceding trees. The model minimizes a regularized objective function, which balances the loss function and a penalty term to reduce model complexity and overfitting. The objective function in XGBoost can be expressed as:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k), where\ \Omega(f) = \gamma T + \frac{1}{2}\lambda||\omega||^2 \tag{1}$$

Here, $l(y_i, \hat{y}_i)$ is a differentiable loss function (e.g., softmax for multi-class classification), $f_k$ represents the individual trees, $T$ is the number of leaves in the tree, and www are the weights on the leaves. The regularization parameters ϒ and λ help control model complexity.

To prepare data for XGBoost, feature extraction is performed on segmented time windows of sensor data, extracting metrics such as mean, variance, and signal magnitude. These features serve as input for the model. Given XGBoost's ability to model non-linear interactions and handle missing data, it is particularly well-suited for time-series sensor inputs, where class imbalance and noise are common. When combined with SMOTE to balance the class distribution, XGBoost is capable of achieving high classification accuracy and generalization, making it an effective model for human activity recognition tasks.

## 3.5. Model Performance Evaluation Methods

To assess the effectiveness of the XGBoost classifier on the mHealth dataset, multiple evaluation techniques are used, each providing a unique perspective on model performance beyond simple accuracy metrics.

One essential tool is the confusion matrix, which provides a detailed breakdown of the classifier's predictions. It shows the number of true positives, true negatives, false positives, and false negatives for each class, making it easier to identify which activities are being misclassified. From the confusion matrix, metrics such as precision, recall, and F1-score can be derived for each class—critical when dealing with imbalanced datasets, as they highlight how well the model performs across all activity types, not just the majority class.

In addition to the confusion matrix, more nuanced visual tools are employed. The Precision-Recall (PR) Curve plots precision against recall for various threshold values, providing insights into the trade-off between false positives and false negatives especially useful for imbalanced classification problems. The Receiver Operating Characteristic (ROC) Curve illustrates the relationship between the true positive rate and false positive rate across thresholds. The area under the ROC curve (AUC-ROC) serves as a summary statistic of the model's discriminatory power. Another valuable method is the Calibration Curve, which compares predicted probabilities with actual outcomes to assess how well the predicted probabilities align with true likelihoods. A well-calibrated model will have a curve that closely follows the diagonal line, indicating reliable probability estimates. Using these methods in combination offers a comprehensive evaluation of the classifier's accuracy, robustness, and reliability in real-world applications.

# 4. Results and Discussion

## 4.1. Result

The evaluation of the classification model without applying SMOTE reveals a performance disparity between the majority and minority classes, primarily due to the imbalance in the dataset. While the overall accuracy reaches 97%, this metric is heavily influenced by the model's success in classifying the dominant class (class 0), which contains the most samples. Class 0 achieves nearly perfect metrics with precision, recall, and F1-score values close to or above 0.98. In contrast, the minority class (class 1) shows weaker performance, with a recall of only 0.75 and an F1-score of 0.7960, indicating the model's difficulty in correctly identifying and predicting this underrepresented activity. The macro average, which gives equal weight to both classes regardless of sample count, is lower than the weighted average, highlighting the skew in class performance. Furthermore, the Area Under the Precision-Recall Curve (AUC-PR) is 0.8100, suggesting that while the model is somewhat capable of distinguishing between classes, its performance on the minority class is suboptimal. These findings are summarized in table 1, which presents the detailed precision, recall, F1-score, support, and AUC-PR for each class and aggregated averages. The results underline the need for balancing techniques like SMOTE to improve minority class recognition.

**Table 1.** Evaluation Metrics of the Model Without SMOTE

| Metric | Class 0 | Class 1 | Macro Avg | Weighted Avg |
|---|---|---|---|---|
| Precision | 0.9800 | 0.8500 | 0.9150 | 0.9680 |
| Recall | 0.9900 | 0.7500 | 0.8700 | 0.9700 |
| F1-score | 0.9850 | 0.7960 | 0.8900 | 0.9680 |
| Support | 4000 | 500 | — | 4500 |
| Accuracy | | | — | 0.9700 |
| AUC-PR | | | — | 0.8100 |

The implementation of SMOTE led to noticeable improvements in the classification model's ability to recognize the minority class in the mHealth dataset. Although the overall accuracy decreased slightly from 97% to 96.5%, this marginal drop is offset by a substantial performance boost in identifying the minority class (class 1). With SMOTE, class 1's precision increased from 0.8500 to 0.8800, and recall improved from 0.7500 to 0.8500, resulting in a higher F1-score of 0.8650. These gains indicate a more balanced and fair classification model that can better distinguish underrepresented activities. Meanwhile, class 0 (majority class) maintained high performance with only a minor reduction in precision and recall, demonstrating that applying SMOTE did not significantly compromise the model's ability to recognize the dominant activity.

Furthermore, the macro average and weighted average metrics both improved compared to the model without SMOTE, reflecting a better overall balance between classes. The macro F1-score increased to 0.9200, showing a more equitable treatment of both activity types, and the weighted average F1-score rose to 0.9640. Notably, the AUC-PR value jumped to 0.8800 from 0.8100, underscoring the model's enhanced precision-recall trade-off, especially for the minority class. These results, detailed in table 2, confirm the effectiveness of SMOTE in mitigating class imbalance and improving classification robustness in human activity recognition tasks.

**Table 2.** Evaluation Metrics of the Model With SMOTE

| Metric | Class 0 | Class 1 | Macro Avg | Weighted Avg |
|---|---|---|---|---|
| Precision | 0.9700 | 0.8800 | 0.9250 | 0.9630 |
| Recall | 0.9800 | 0.8500 | 0.9150 | 0.9650 |
| F1-score | 0.9750 | 0.8650 | 0.9200 | 0.9640 |
| Support | 4000 | 500 | — | 4500 |
| Accuracy | | | — | 0.9650 |
| AUC-PR | | | — | 0.8800 |

The confusion matrices presented in Figure 2 illustrate the classification results for models trained without and with the application of SMOTE on the mHealth dataset. The left matrix reflects the model's performance without SMOTE, where 3934 out of 4000 majority class (class 0) instances were correctly classified, while 66 were misclassified as the minority class (class 1). For class 1, only 375 out of 500 instances were correctly predicted, and 125 were misclassified as class 0, indicating a relatively high false negative rate and a model bias toward the majority class due to the data imbalance.

After applying SMOTE, the right matrix shows notable improvements in minority class detection. The number of correctly classified class 1 instances increased to 425, with false negatives reduced to 75. Meanwhile, the model continued to perform strongly on class 0, correctly identifying 3943 instances and slightly increasing the misclassified count to 57. This visual evidence confirms that SMOTE improves recall and overall recognition of the minority class while maintaining high accuracy for the majority class. As shown in figure 2, SMOTE effectively addresses class imbalance, resulting in a more robust and equitable classification model for human activity recognition.

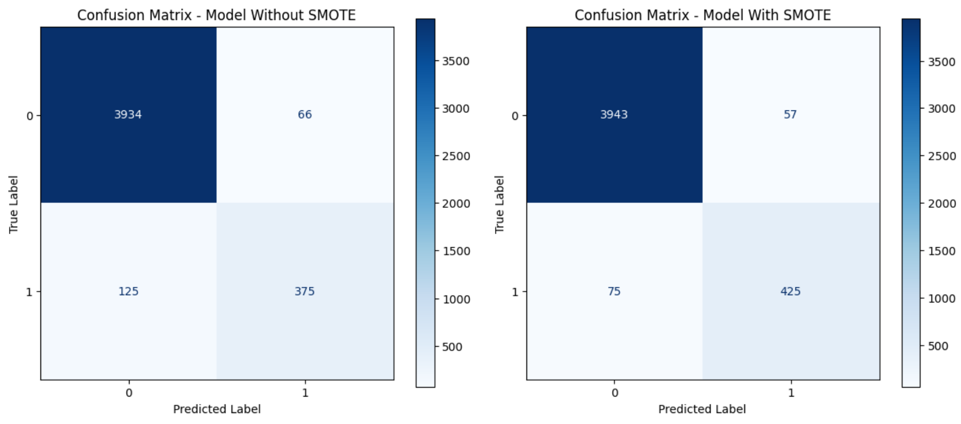


**Figure 2.** Confusion Matrix Comparison Between Models with SMOTE and Without SMOTE

The comparison of the ROC curves in figure 3 demonstrates the classification performance of the models trained without and with SMOTE in terms of their ability to distinguish between classes. The ROC curve plots the true positive rate (sensitivity) against the false positive rate at various threshold levels, offering insight into the trade-off between detecting positive cases and avoiding false alarms.

The model without SMOTE achieves an Area Under the Curve (AUC) of 0.95, while the model with SMOTE slightly outperforms it with an AUC of 0.96. This improvement, though modest, indicates that the model trained with SMOTE is better at differentiating between the positive (minority) and negative (majority) classes across all decision thresholds.

ISSN 2776-8007


The closer the ROC curve is to the top-left corner of the graph, the better the model performance. The result supports previous evaluations, showing that SMOTE not only improves recall for the minority class but also enhances the model's overall discriminative capability, making it more reliable in scenarios with class imbalance.

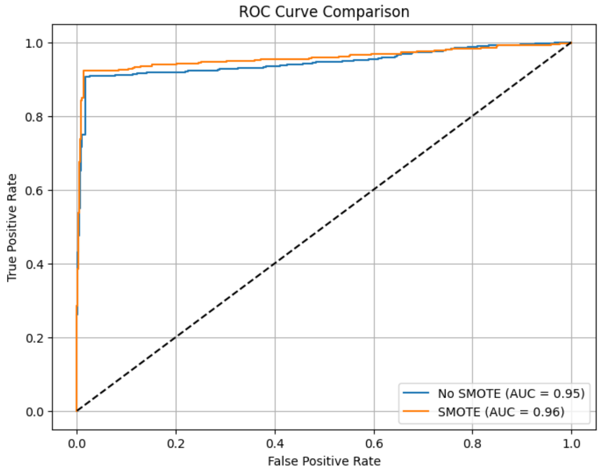


**Figure 3.** ROC Curve Comparison Between Models with SMOTE and Without SMOTE

The PR curves displayed in figure 4 illustrate the performance comparison between the models trained without and with SMOTE, focusing on their ability to identify the minority class in the mHealth dataset. Unlike the ROC curve, which considers both true positive and false positive rates, the PR curve provides a more informative evaluation in imbalanced classification settings by plotting precision (positive predictive value) against recall (sensitivity).

As shown, the model with SMOTE consistently maintains higher precision across varying recall levels, resulting in a superior AP score of 0.91 compared to 0.88 for the model without SMOTE. This improvement signifies that the SMOTE-enhanced model is more reliable in correctly identifying true instances of the minority class without sacrificing accuracy. The smoother and more stable curve also indicates better generalization, especially in scenarios where the cost of false positives and false negatives is high. These findings support the conclusion that SMOTE enhances the model’s predictive quality for underrepresented classes in imbalanced datasets.

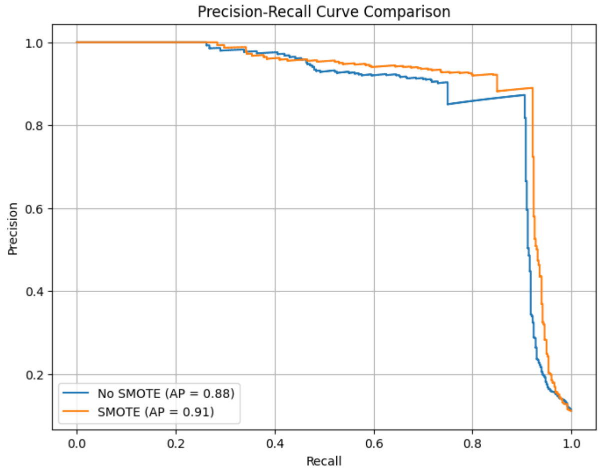


**Figure 4.** Precision-Recall Curve Comparison Between Models with SMOTE and Without SMOTE

The plot illustrates how precision, recall, and F1 score vary as the classification threshold changes. At low threshold values, recall is high because the model classifies most instances as positive, but precision is low due to many false positives. As the threshold increases, precision improves because the model becomes more selective, while recall decreases since fewer positive instances are identified. The F1 score, which represents the harmonic mean of precision and recall, peaks around a threshold of 0.25, indicating the optimal balance between correctly identifying positive cases and minimizing false alarms. This analysis is crucial for determining the most appropriate threshold for the model depending on whether precision or recall is prioritized in the application. These insights are visually summarized in the curve shown in figure 5.

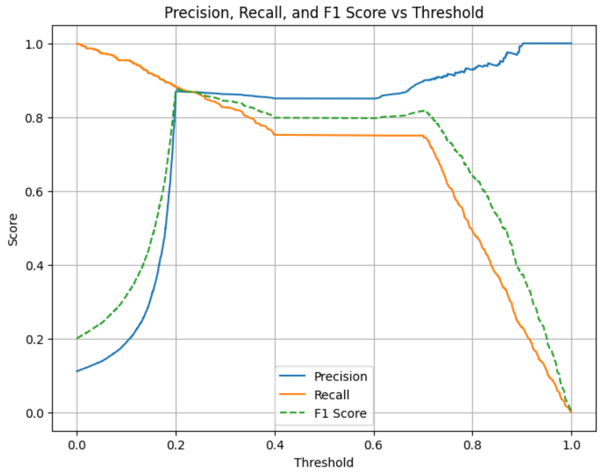


**Figure 5**. Precision, Recall, and F1 Score Across Different Thresholds

The calibration curve in figure 6 assesses how well the predicted probabilities from the classification model correspond to the actual observed outcomes. Ideally, a perfectly calibrated model would have its predictions lie along the diagonal dashed line, indicating that the predicted probability matches the true fraction of positive cases. In this case, the model's calibration curve deviates from the ideal line at several points, revealing discrepancies between predicted probabilities and actual event frequencies. These deviations suggest instances of both overconfidence and under confidence in the model's probability estimates. Although the model performs well in classification, this indicates that its probability outputs may require further calibration to be trustworthy for applications relying on accurate risk assessment or probability-based decision-making.

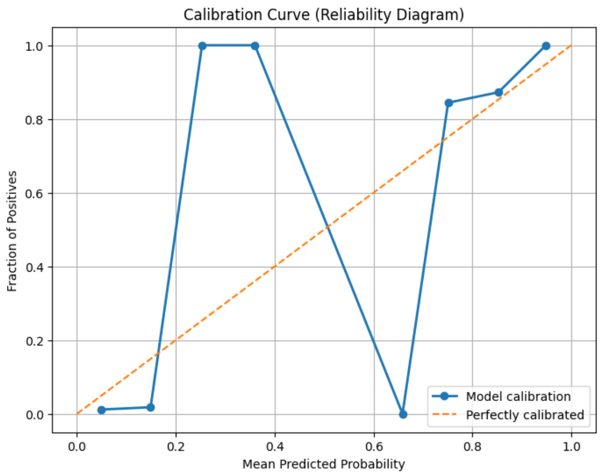


**Figure 6.** Calibration Curve of the Classification Model

## 4.2. Discussion

The present study investigates the critical issue of class imbalance in wearable sensor-based HAR using the mHealth dataset. Our findings reinforce the well-known challenge that imbalanced datasets significantly hinder model performance, particularly for minority classes that often represent critical but infrequent activities. The baseline model trained without addressing class imbalance yielded an overall accuracy of 97%, yet this metric masked the model's underperformance on the minority class (class 1), where recall dropped to 0.75 and the F1-score to 0.7960. This performance disparity clearly illustrates that high accuracy in imbalanced datasets can be misleading, as it predominantly reflects the model's ability to classify the majority class accurately. This aligns with Kumar's [30] argument emphasizing that accuracy alone is insufficient to assess models on skewed datasets and highlights the need for more nuanced evaluation metrics such as recall, precision, F1-score, and area under the Precision-Recall Curve (AUC-PR).

The imbalance problem is a pervasive obstacle in HAR research, as documented by Zhang et al. [26], who detailed the negative impact of class imbalance on recognition accuracy, especially for rare activities that tend to have fewer training samples. Our results echo these observations and extend them by quantifying the extent to which imbalance skews classifier performance in the context of the mHealth dataset. The lower macro-average scores further confirm that models trained on imbalanced data disproportionately favor the majority class, leading to poor generalization on underrepresented activity types. This phenomenon is exacerbated by class overlap, where feature distributions of different classes intersect, causing further confusion for classifiers, as discussed by Lv and Liu [27].

To counteract this challenge, we applied the SMOTE, a widely adopted method that synthesizes new minority class examples by interpolating between existing samples. The application of SMOTE resulted in substantial improvements in minority class recognition: precision rose from 0.85 to 0.88, recall improved markedly from 0.75 to 0.85, and the F1-score increased to 0.8650. Although the overall accuracy saw a slight decrease from 97% to 96.5%, this trade-off is acceptable given the significant gains in minority class detection, which is often more critical in real-world HAR applications where missing rare but important activities can have serious consequences. The increase in AUC-PR from 0.81 to 0.88 further validates the enhanced ability of the SMOTE-augmented model to distinguish between classes across various decision thresholds.

These results support previous work by Elreedy et al. [31], who demonstrated the effectiveness of SMOTE in improving classifier sensitivity and robustness in imbalanced scenarios, including wearable sensor datasets. Our confusion matrix analysis visually corroborates these findings by showing fewer false negatives for the minority class post-SMOTE, and ROC curve comparisons reveal a modest but meaningful improvement in the model's discriminative power with SMOTE, raising the AUC from 0.95 to 0.96. These findings collectively suggest that SMOTE is a viable and effective strategy for balancing datasets in HAR, consistent with Hartono and Ongko's [32] advocacy for hybrid approaches that combine oversampling with ensemble learning to improve predictive performance in multi-class imbalanced datasets.

Further, the Precision-Recall curves illustrate that the model with SMOTE consistently outperforms the baseline across a wide range of recall values, achieving a higher average precision (AP) of 0.91 versus 0.88 without SMOTE. This indicates that the SMOTE-enhanced model maintains higher precision without sacrificing recall, a crucial factor when minimizing false positives and negatives is essential. Threshold analysis further clarifies the trade-offs between precision and recall at different decision thresholds, with the optimal F1-score attained around a threshold of 0.25. This insight is valuable for practical deployment, where application-specific priorities may dictate the choice of threshold to balance sensitivity and specificity effectively.

Despite these encouraging improvements, the calibration curve indicates that the predicted probabilities from the SMOTE-trained model are not perfectly calibrated. The deviation from the ideal diagonal line suggests the presence of overconfident and underconfident probability estimates. This lack of calibration can undermine the model's reliability in scenarios where probability scores inform critical decision-making, such as healthcare monitoring or emergency interventions. Calibration issues are a known limitation in many machine learning models, including those trained with oversampling techniques, and highlight the necessity of incorporating calibration methods such as Platt scaling or isotonic regression in future work.

This study advances the field of wearable sensor-based human activity recognition by demonstrating how the SMOTE can effectively address the pervasive challenge of class imbalance in the mHealth dataset. Unlike many recent works that rely heavily on deep learning, this research shows that classical machine learning models like XGBoost, when paired with appropriate data balancing methods, can achieve robust and balanced classification performance. The comprehensive evaluation including metrics such as precision, recall, F1-score, and visual tools like ROC and Precision-Recall curves provides a detailed understanding of how SMOTE enhances minority class recognition without significantly harming majority class accuracy. This practical insight is particularly valuable for real-world applications where computational resources or dataset sizes may limit the feasibility of more complex models.

The implications of these findings are significant for designing wearable monitoring systems in healthcare, elder care, and rehabilitation, where detecting rare but critical activities is crucial. Improved minority class detection ensures that important but infrequent events, such as falls or unusual physiological patterns, are less likely to be missed, potentially enhancing user safety and care outcomes. Moreover, the ability to tune classification thresholds based on precision-recall trade-offs allows system designers to customize models according to specific needs, balancing sensitivity and specificity as appropriate for the application context. The study also highlights the importance of model calibration, which affects how trustworthy the predicted probabilities are for decision making a critical consideration in clinical settings where risk estimates guide interventions.

Despite these advances, limitations remain. The generation of synthetic minority samples through SMOTE can sometimes introduce noise or increase class overlap, potentially complicating classification and reducing model reliability in some cases. This underscores the need for careful implementation, possibly through hybrid oversampling and data cleaning methods to optimize results. Additionally, while the current work focuses on a binary classification framework, real-world activity recognition often involves multiple imbalanced classes, requiring further exploration of SMOTE's effectiveness in multi class scenarios. Lastly, the imperfect calibration observed suggests that future research should incorporate probability calibration techniques to improve confidence in model outputs. Integrating these improvements with more advanced sensor fusion and deep learning approaches may further enhance the accuracy and applicability of wearable sensor-based activity recognition systems.

**5. Conclusion**

This study addresses the persistent challenge of class imbalance in wearable sensor-based human activity recognition using the mHealth dataset. Although the baseline XGBoost classifier achieved high overall accuracy of 97%, its performance on the minority class was notably poor, with a recall of 0.75 and an F1-score of 0.796. This discrepancy highlights the limitations of relying solely on accuracy as a performance metric in imbalanced datasets, as it often masks the model's inability to correctly classify rare but important activities. In healthcare and related fields, detecting these minority class activities is critical for timely intervention and accurate monitoring.

By applying the SMOTE, we generated synthetic samples to augment the minority class, leading to significant improvements in model performance. The recall and F1-score for the minority class increased substantially to 0.85 and 0.865, respectively, while maintaining strong performance on the majority class. Although the overall accuracy dropped slightly to 96.5%, this trade-off resulted in a more balanced and fairer classifier. Additionally, the AUC-PR improved from 0.81 to 0.88, indicating enhanced ability to discriminate between classes under imbalanced conditions. These findings align with previous research emphasizing the benefits of SMOTE for imbalanced classification problems and demonstrate its applicability to wearable sensor data.

Despite these improvements, the calibration analysis revealed that the predicted probabilities are not perfectly reliable, showing both overconfidence and under confidence at certain thresholds. This limitation suggests that further work is needed to calibrate model outputs, potentially through methods like Platt scaling or isotonic regression, to improve the trustworthiness of probability estimates in real-world applications. Overall, this study confirms that combining SMOTE with robust classifiers such as XGBoost can effectively mitigate class imbalance in activity recognition, enhancing the detection of rare activities critical for healthcare and sports monitoring. Future research should explore extensions to multi-class imbalanced datasets and the integration of probability calibration techniques for improved practical utility.

**6. Declarations**

## 6.1. Author Contributions

Conceptualization: S., C.R.A.W., I.R.Y.; Methodology: S., I.R.Y.; Software: C.R.A.W.; Validation: S., I.R.Y.; Formal Analysis: S.; Investigation: C.R.A.W.; Resources: I.R.Y.; Data Curation: C.R.A.W.; Writing – Original Draft Preparation: S.; Writing – Review and Editing: S., C.R.A.W., I.R.Y.; Visualization: C.R.A.W.; All authors have read and agreed to the published version of the manuscript.

## 6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

## 6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

## 6.4. Institutional Review Board Statement

Not applicable.

## 6.5. Informed Consent Statement

Not applicable.

## 6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**References**

[1] S. B. Junaid, A. A. Imam, A. O. Balogun, L. C. De Silva, Y. A. Surakat, G. Kumar, M. Abdulkarim, A. N. Shuaibu, A. Garba, Y. Sahalu, A. Mohammed, T. Y. Mohammed, B. A. Abdulkadir, A. A. Abba, N. A. I, Kakumi, and S. Mahamad, "Recent Advancements in Emerging Technologies for Healthcare Management Systems: A Survey," *Healthcare*, vol. 10, no. 10, pp. 1940, 2022, doi: 10.3390/healthcare10101940.

[2] A. Ç. Seçkin, B. Ateş, and M. Seçkin, "Review on Wearable Technology in Sports: Concepts, Challenges and Opportunities," *Appl. Sci.*, vol. 13, no. 18, pp. 10399, 2023, doi: 10.3390/app131810399.

[3] Z. Zhao, Q. Li, Y. Dong, J. Gong, Z. Li, and J. Zhang, "Washable Patches With Gold Nanowires/Textiles in Wearable Sensors for Health Monitoring," *ACS Appl. Mater. Interfaces*, vol. 14, no. 16, pp. 18884–18900, 2022, doi: 10.1021/acsami.2c01729.

[4] M. Akter, H. R. Anik, S. I. Tushar, I. S. Tania, Md. K. H. Chowdhury, S. Md. M. Hasan, and B. F. Bristy, "Advances in Functionalized Applications of Graphene-Based Wearable Sensors in Healthcare," *Adv. Sens. Res.*, vol. 3, no. 2, pp. 1–31, 2023, doi: 10.1002/adsr.202300120.

[5] A. S. Barkah, S. R. Selamat, Z. Z. Abidin, and R. Wahyudi, "Impact of Data Balancing and Feature Selection on Machine Learning-Based Network Intrusion Detection," *Int. J. Informatics Vis.*, vol. 7, no. 1, pp. 241–248, 2023, doi: 10.30630/joiv.7.1.1041.

[6] F. Alharbi, L. Ouarbya, and J. A. Ward, "Comparing Sampling Strategies for Tackling Imbalanced Data in Human Activity Recognition," *Sensors*, vol. 22, no. 4, pp. 1373, 2022, doi: 10.3390/s22041373.

[7] S. Stiller, J. F. Dueñas, S. Hempel, M. C. Rillig, and M. Ryo, "Deep Learning Image Analysis for Filamentous Fungi Taxonomic Classification: Dealing With Small Datasets With Class Imbalance and Hierarchical Grouping," *Biol. Methods Protoc.*, vol. 9, no. 1, pp. 1–11, 2024, doi: 10.1093/biomethods/bpae063.

[8] S. Y. Alaba, M. M. Nabi, C. Shah, J. Prior, M. D. Campbell, F. Wallace, J. E. Ball, and R. Moorhead, "Class-Aware Fish Species Recognition Using Deep Learning for an Imbalanced Dataset," *Sensors*, vol. 22, no. 21, pp. 8268, 2022, doi: 10.3390/s22218268.

[9] M. E. Sánchez-Gutiérrez and P. P. González-Pérez, "Addressing the Class Imbalance in Tabular Datasets From a Generative Adversarial Network Approach in Supervised Machine Learning," *J. Algorithm. Comput. Technol.*, vol. 17, no. November, pp. 1–17, 2023, doi: 10.1177/17483026231215186.

[10] H. I. Naufal, A. Efendi, and E. Sumarminingsih, "Bayesian Additive Regression Trees for Classification of Unbalanced Class of Credit Collectability Data," *Asian J. Probab. Stat.*, vol. 23, no. 1, pp. 16–27, 2023, doi: 10.9734/ajpas/2023/v23i1494.

[11] R. Saritha and R. Gunasundari, "Enhanced Transformer-Based Deep Kernel Fused Self Attention Model for Lung Nodule Segmentation and Classification," *Arch. Tech. Sci.*, vol. 31, no. 2, pp. 175–191, 2024, doi: 10.70102/afts.2024.1631.175.

[12] D. K. Verma, G. Saxena, A. Paraye, A. Rajan, A. Rawat, and R. K. Verma, "Classifying COVID-19 and Viral Pneumonia Lung Infections Through Deep Convolutional Neural Network Model Using Chest X-Ray Images," *J. Med. Phys.*, vol. 47, no. 1, pp. 57–64, 2022, doi: 10.4103/jmp.jmp_100_21.

[13] D. Na, J. Kim, K. Chung, S. Han, G. Yun, J. H. Choi, H. C. Kim, Y. Cho, and J. P. Jeon, "Classification of Carotid Plaque

Vulnerability by Neurosurgical Residents Using Ultrasonography in the Clinical Field," *J. Korean Soc. Geriatr. Neurosurg.*, vol. 18, no. 2, pp. 38–43, 2022, doi: 10.51638/jksgn.2022.00094.

[14] B. K. Jang and Y. R. Park, "Development and Validation of Adaptable Skin Cancer Classification System Using Dynamically Expandable Representation," *Healthc. Inform. Res.*, vol. 30, no. 2, pp. 140–146, 2024, doi: 10.4258/hir.2024.30.2.140.

[15] K. Leon-López, F. Mouret, H. Argüello, and J. Tourneret, "Anomaly Detection and Classification in Multispectral Time Series Based on Hidden Markov Models," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. August, pp. 1–11, 2022, doi: 10.1109/tgrs.2021.3101127.

[16] B. Li, L. Hu, S. Zhang, S. Li, W. Tang, and G. Chen, "The Application and Clinical Translation of the Self-Evolving Machine Learning Methods in Predicting Diabetic Retinopathy and Visualizing Clinical Transformation," *Front. Endocrinol. (Lausanne).*, vol. 15, no. September, pp. 1–11, 2024, doi: 10.3389/fendo.2024.1429974.

[17] M. Kim and K. Hwang, "An Empirical Evaluation of Sampling Methods for the Classification of Imbalanced Data," *PLoS One*, vol. 17, no. 7, pp. e0271260, 2022, doi: 10.1371/journal.pone.0271260.

[18] P. Sa-nguannarm, E. Elbasani, and J. Kim, "Human Activity Recognition for Analyzing Stress Behavior Based on Bi-LSTM," *Technol. Heal. Care*, vol. 31, no. 5, pp. 1997–2007, 2023, doi: 10.3233/thc-235002.

[19] S. K. Yadav, A. Luthra, K. Tiwari, H. M. Pandey, and S. Akbar, "ARFDNet: An Efficient Activity Recognition Fall Detection System Using Latent Feature Pooling," *Knowledge-Based Syst.*, vol. 239, no. March, pp. 107948, 2022, doi: 10.1016/j.knosys.2021.107948.

[20] B. M. V Guerra, E. Torti, E. Marenzi, M. Schmid, S. Ramat, F. Leporati, and G. Danese, "Ambient Assisted Living for Frail People Through Human Activity Recognition: State of the Art, Challenges and Future Directions," *Front. Neurosci.*, vol. 17, no. October, pp. 1–18, 2023, doi: 10.3389/fnins.2023.1256682.

[21] A. Nouriani, A. Jonason, L. T. Sabal, J. T. Hanson, J. N. Jean, T. Lisko, E. Reid, Y. Moua, S. Rozeboom, K. Neverman, C. Stowe, R. Rajamani, and R. A. McGovern, "Real World Validation of Activity Recognition Algorithm and Development of Novel Behavioral Biomarkers of Falls in Aged Control and Movement Disorder Patients," *Front. Aging Neurosci.*, vol. 15, no. February, pp. 1–10, 2023, doi: 10.3389/fnagi.2023.1117802.

[22] E. F. Asl, S. Ebadollahi, R. Vahidnia, and A. Jalali, "Statistical Database of Human Motion Recognition Using Wearable IoT: A Review," *IEEE Sens. J.*, vol. 23, no. 14, pp. 15253–15304, 2023, doi: 10.1109/jsen.2023.3282171.

[23] R. Brard, L. Bellanger, L. Chevreuil, F. Doistau, P. Drouin, and A. Stamm, "A Novel Walking Activity Recognition Model for Rotation Time Series Collected by a Wearable Sensor in a Free-Living Environment," *Sensors*, vol. 22, no. 9, pp. 3555, 2022, doi: 10.3390/s22093555.

[24] H. Ramirez, S. A. Velastín, S. Cuellar-Carrillo, E. Fábregas, and G. Farías, "BERT for Activity Recognition Using Sequences of Skeleton Features and Data Augmentation With GAN," *Sensors*, vol. 23, no. 3, pp. 1400, 2023, doi: 10.3390/s23031400.

[25] A. Vijayvargiya, B. Singh, R. Kumar, U. Desai, and D. J. Hemanth, "Hybrid Deep Learning Approaches for sEMG Signal-Based Lower Limb Activity Recognition," *Math. Probl. Eng.*, vol. 2022, no. 1, pp. 1–12, 2022, doi: 10.1155/2022/3321810.

[26] J. Zhang, Y. Li, B. Liu, H. Chen, J. Zhou, H. Yu, and B. Qin, "A Broad TSK Fuzzy Classifier With a Simplified Set of Fuzzy Rules for Class-Imbalanced Learning," *Mathematics*, vol. 11, no. 20, pp. 4284, 2023, doi: 10.3390/math11204284.

[27] Z. Lv and Q. Liu, "Imbalanced Data Over-Sampling Method Based on ISODATA Clustering," *Ieice Trans. Inf. Syst.*, vol. E106.D, no. 9, pp. 1528–1536, 2023, doi: 10.1587/transinf.2022edp7190.

[28] H. S. Bako, F. U. Ambursa, B. S. Galadanci, and M. Garba, "Predicting Timely Graduation of Postgraduate Students Using Random Forests Ensemble Method," *Fudma J. Sci.*, vol. 7, no. 3, pp. 177–185, 2023, doi: 10.33003/fjs-2023-0703-1773.

[29] M. Hosseini, A. H. Rasekh, and A. Keshavarzi, "Improving Clinical Abbreviation Sense Disambiguation Using Attention-based Bi-LSTM and Hybrid Balancing Techniques in Imbalanced Datasets," *J. Eval. Clin. Pract.*, vol. 30, no. 7, pp. 1327–1336, 2024, doi: 10.1111/jep.14041.

[30] A. Kumar, "SOM-US: A Novel Under-Sampling Technique for Handling Class Imbalance Problem," *J. Commun. Softw. Syst.*, vol. 20, no. 1, pp. 69–75, 2024, doi: 10.24138/jcomss-2023-0133.

[31] D. Elreedy, A. F. Atiya, and F. Kamalov, "A Theoretical Distribution Analysis of Synthetic Minority Oversampling Technique (SMOTE) for Imbalanced Learning," *Mach. Learn.*, vol. 113, no. 7, pp. 4903–4923, 2023, doi: 10.1007/s10994-022-06296-4.

[32] H. Hartono and E. Ongko, "Hybrid Approach With Distance Feature for Multi-Class Imbalanced Datasets," *Int. J. Informatics Vis.*, vol. 7, no. 1, pp. 131–138, 2023, doi: 10.30630/joiv.7.1.1292.

[33] G. M. Husain, D. Nasef, R. Jose, J. Mayer, M. Bekbolatova, T. Devine, and M. Toma, "SMOTE vs. SMOTEENN: A Study

on the Performance of Resampling Algorithms for Addressing Class Imbalance in Regression Models," *Algorithms*, vol. 18, no. 1, pp. 37, 2025, doi: 10.3390/a18010037.

[34] B. Zhou, D. Geissler, M. Fauhaber, C. E. Gleiss, E. F. Zahn, L. S. S. Ray, D. Gamarra, V. F. Rey, S. Suh, S. Bian, G. Joost, and P. Lukowicz, "MoCaPose: Motion Capturing with Textile-integrated Capacitive Sensors in Loose-fitting Smart Garments," *Proc. Acm Interact. Mob. Wearable Ubiquitous Technol.*, vol. 7, no. 1, pp. 1–40, 2023, doi: 10.1145/3580883.

[35] M. K. Hooshmand and D. Hosahalli, "Network Anomaly Detection Using Deep Learning Techniques," *CAAI Trans. Intell. Technol.*, vol. 7, no. 2, pp. 228–243, 2022, doi: 10.1049/cit2.12078.

[36] S. Wang, L. Ma, and J. Wang, "Fault Diagnosis Method Based on CND-SMOTE and BA-SVM Algorithm," *J. Phys. Conf. Ser.*, vol. 2493, no. 1, pp. 12008, 2023, doi: 10.1088/1742-6596/2493/1/012008.

[37] H. Chen, J. Liu, J. Wang, and W. Shi, "DAPLSR: Data Augmentation Partial Least Squares Regression Model via Manifold Optimization," *Int. J. Cybern. Informatics*, vol. 13, no. 5, pp. 101–118, 2024, doi: 10.5121/ijci.2024.130507.

[38] S. D. Abdullahi and S. A. Muhammad, "Early Prediction of Cerebrovascular Disease Using Boosting Machine Learning Algorithms to Assist Clinicians," *J. Appl. Sci. Environ. Manag.*, vol. 26, no. 6, pp. 1031–1037, 2022, doi: 10.4314/jasem.v26i6.6.

[39] L. Duan, W. Xue, J. Huang, and X. Zheng, "Joint Sample Position Based Noise Filtering and Mean Shift Clustering for Imbalanced Classification Learning," *Tsinghua Sci. Technol.*, vol. 29, no. 1, pp. 216–231, 2024, doi: 10.26599/tst.2023.9010006.