mathematics

MDPI

Article

High-Performance Real-Time Human Activity Recognition Using Machine Learning

Pardhu Thottempudi [1,2,3], Biswaranjan Acharya [4,*] and Fernando Moreira [5,6,*]

[1] Department of Electronics and Communications Engineering, BVRIT HYDERABAD College of Engineering for Women, Hyderabad 500090, India; pardhu.t@bvrithyderabad.edu.in or pardhu@unimap.edu.my
[2] Faculty of Electronic Engineering & Technology, Universiti Malaysia Perlis, Pauh Putra Campus, Arau 02600, Perlis, Malaysia
[3] Centre of Excellence for Micro System Technology (MiCTEC), Universiti Malaysia Perlis, Pauh Putra Campus, Arau 02600, Perlis, Malaysia
[4] Department of Computer Engineering-AI & BDA, Marwadi University, Rajkot 360003, India
[5] REMIT, IJP, Universidade Portucalense, 4200 Porto, Portugal
[6] IEETA, Universidade de Aveiro, 3810 Aveiro, Portugal
* Correspondence: biswaranjan.acharya@marwadieducation.edu.in (B.A.); fmoreira@upt.pt (F.M.)

Abstract: Human Activity Recognition (HAR) is a vital technology in domains such as healthcare, fitness, and smart environments. This paper presents an innovative HAR system that leverages machine-learning algorithms deployed on the B-L475E-IOT01A Discovery Kit, a highly efficient microcontroller platform designed for low-power, real-time applications. The system utilizes wearable sensors (accelerometers and gyroscopes) integrated with the kit to enable seamless data acquisition and processing. Our model achieves outstanding performance in classifying dynamic activities, including walking, walking upstairs, and walking downstairs, with high precision and recall, demonstrating its reliability and robustness. However, distinguishing between static activities, such as sitting and standing, remains a challenge, with the model showing a lower recall for sitting due to subtle postural differences. To address these limitations, we implement advanced feature extraction, data augmentation, and sensor fusion techniques, which significantly improve classification accuracy. The ease of use of the B-L475E-IOT01A kit allows for real-time activity classification, validated through the Tera Term interface, making the system ideal for practical applications in wearable devices and embedded systems. The novelty of our approach lies in the seamless integration of real-time processing capabilities with advanced machine-learning techniques, providing immediate, actionable insights. With an overall classification accuracy of 90%, this system demonstrates great potential for deployment in health monitoring, fitness tracking, and eldercare applications. Future work will focus on enhancing the system's performance in distinguishing static activities and broadening its real-world applicability.

Keywords: human activity recognition; machine learning; wearable sensors; real-time classification; feature extraction; tera term; sensor fusion; health monitoring; fitness tracking; deep learning

MSC: 68T05; 68U35

check for updates

Citation: Thottempudi, P.; Acharya, B.; Moreira, F. High-Performance Real-Time Human Activity Recognition Using Machine Learning. *Mathematics* **2024**, *12*, 3622. https://doi.org/10.3390/math12223622

Academic Editors: Qingshan Jiang and Huawen Liu

Received: 30 September 2024
Revised: 28 October 2024
Accepted: 12 November 2024
Published: 20 November 2024

1. Introduction

Human activity recognition (HAR) has become a crucial component in the development of intelligent systems, with applications spanning from health monitoring and fitness tracking to eldercare and smart environments [1–4]. The ability to automatically classify and monitor physical activities has enabled new opportunities for providing personalized feedback in real time, facilitating the early detection of health issues, and improving the overall quality of life for users. This growing interest in HAR is driven by the increasing availability of wearable sensors and advancements in machine-learning algorithms that can process sensor data efficiently [5–7].

Wearable devices such as smartphones, smartwatches, and dedicated fitness trackers have emerged as prominent platforms for HAR systems due to their widespread adoption and integration of advanced sensors like accelerometers, gyroscopes, and magnetometers [6,8]. These sensors capture continuous streams of data reflecting the user's movement patterns, which can be leveraged to classify activities such as walking, running, sitting, and more complex motion sequences like stair climbing or descending [4]. Despite advancements in sensor technology, distinguishing between similar postural activities—such as sitting and standing—remains a challenging task, as these activities exhibit subtle variations that are not easily captured by conventional sensor setups [9].

In recent years, the integration of machine-learning techniques into HAR has significantly improved the accuracy and reliability of activity classification models. Deep learning, in particular, has shown great promise in extracting meaningful patterns from raw sensor data without the need for manual feature engineering [6]. Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and hybrid models have been applied successfully to recognize complex human activities [7,10]. However, while these methods perform well in classifying dynamic activities like walking or running, challenges persist in the classification of static or transition activities, which require more precise feature extraction and data augmentation techniques.

While our model demonstrates high accuracy in activity recognition, it is essential to clarify the testing conditions under which these results were obtained. The controlled lab environment minimized variables such as lighting conditions, ambient noise, and movement interferences, which can impact sensor performance. In practical applications, such as home environments or outdoor settings, factors including sensor misalignment, variable lighting, and non-standardized user movement patterns may introduce inconsistencies, impacting classification accuracy. Addressing these practical limitations is critical to deploying HAR systems effectively beyond the lab.

The current study aims to address some of these challenges by proposing a robust HAR system that utilizes a combination of sensors and machine-learning models to classify both dynamic and static activities. Our system integrates data from wearable sensors and processes it using advanced machine-learning models, including real-time classification capabilities demonstrated through Tera Term. The primary objective of this work is to enhance activity recognition accuracy, particularly for activities with subtle distinctions, such as sitting and standing, and to explore the potential of deploying such a system in practical applications like health monitoring, fitness tracking, and eldercare.

This paper is organized as follows: Section 2 details the related work and background in HAR systems, while Section 3 describes the research gaps, Section 4 describes the methodology, including sensor setup, data acquisition, and model architecture. Section 5 presents the experimental results, followed by a discussion of the findings in Section 6. Finally, Section 7 concludes the paper with suggestions for future research directions.

2. Literature Survey

Human Activity Recognition (HAR) has emerged as a critical area of research due to its wide applications in fields such as healthcare, smart homes, fitness tracking, and human-computer interaction (HCI). HAR systems utilize a variety of sensors to monitor and classify human activities, facilitating real-time recognition of behavior patterns. Recent advancements in embedded systems, sensor fusion, and machine learning, especially Tiny Machine Learning (TinyML), have enabled HAR to be deployed on resource-constrained devices, ensuring low-latency, real-time inference directly at the edge where data are collected [11].

2.1. Early Contributions and Body-Worn Sensors

In early work, Bulling et al. [1] provided a comprehensive tutorial on using body-worn inertial sensors for HAR, which laid the groundwork for subsequent innovations. Lara and Labrador [3] conducted a survey of HAR systems, focusing on the use of wearable sensors such as accelerometers, gyroscopes, and magnetometers. These sensors capture human motion data, which is then used to classify activities using various machine-learning techniques. Kwapisz et al. [4] also contributed to the field by demonstrating the potential of cell phone accelerometers for activity recognition in real-world environments.

2.2. Multimodal Sensors and Context Recognition

Recent advancements in HAR have extended to multimodal sensors that combine accelerometers, gyroscopes, and environmental sensors to capture more detailed information. Vaizman et al. [2] explored context recognition using multimodal sensors in real-world scenarios. This approach allows for more accurate and robust activity recognition, especially in uncontrolled environments where noise and variability in human behavior are high.

2.3. Deep Learning and HAR

The application of deep-learning models, particularly Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, has significantly improved the accuracy of HAR systems. Ordoñez and Roggen [7] combined CNNs with LSTM models to effectively handle multimodal sensor data, capturing both spatial and temporal patterns in human activity. This approach marked a turning point in HAR, leading to substantial improvements in classification accuracy. LSTMs, in particular, excel at modeling sequential data, making them highly suitable for time-series data from wearable sensors [12].

Ignatov [5] applied CNNs to real-time HAR using accelerometer data, demonstrating the potential of deep learning to significantly outperform traditional methods in activity classification. Similarly, Wang et al. [6] provided a comprehensive survey of deep-learning techniques for sensor-based activity recognition, outlining the benefits of deep architectures for feature extraction.

2.4. Kinect Sensor for Human Motion Capture

The Microsoft Kinect sensor has played a pivotal role in HAR, particularly in its ability to track skeletal movements in real time. Shotton et al. [13] introduced a system that uses Kinect's depth-sensing capabilities for human pose recognition, offering a precise method for capturing body movements in three dimensions. Kinect's ability to track joints and limbs in real time has made it a popular choice for HAR applications in fields like healthcare and sports [14].

2.5. Preprocessing and Feature Extraction

The preprocessing of sensor data is a critical component in HAR systems, as it reduces noise and ensures the data are suitable for machine-learning models. Shotton et al. [13] and Huang et al. [14] highlighted the importance of normalizing and filtering skeletal data from Kinect to enhance its quality. This preprocessing transforms raw data into structured feature vectors that capture important aspects of human movement, such as joint positions, velocities, and trajectories.

2.6. Machine-Learning Models for HAR

Ravi et al. [8] were among the pioneers in applying machine learning to HAR, utilizing accelerometer data to classify activities like walking, running, and sitting. Their work was built upon by later research that employed deep-learning models to process more complex activity data. Ha et al. [10] proposed multimodal CNNs for activity recognition, demonstrating that integrating different sensor modalities can significantly enhance classification accuracy.

Recent work by Jamil et al. [15] applied AutoML techniques and weighted soft voting ensembles for smartphone-based HAR, optimizing feature extraction and improving overall classification performance. Diraco et al. [16] reviewed multimodal sensor fusion techniques, emphasizing their role in improving HAR system reliability and performance, especially in smart living environments.

2.7. TinyML and Real-Time HAR on Embedded Systems

With the development of TinyML, HAR systems can now be deployed on microcontrollers and other resource-constrained devices. STMicroelectronics' STM32Cube.AI [17] facilitates the deployment of machine-learning models on STM32 microcontrollers, enabling real-time HAR at the edge. This is crucial for applications where low latency and energy efficiency are required, such as wearable fitness trackers and healthcare devices.

David et al. [18] demonstrated the effectiveness of TensorFlow Lite Micro for deploying machine-learning models on embedded systems. Their work highlights the trade-offs between model complexity and resource usage, ensuring that models can run efficiently on low-power devices. Similarly, Kopparapu et al. [19] introduced TinyFedTL, a federated learning system designed for TinyML devices, enabling distributed learning across multiple embedded systems without sharing raw data.

2.8. Federated Learning and Privacy-Preserving HAR

Federated Learning (FL) enables multiple devices to collaboratively train a global model without transmitting raw data, which is especially important in privacy-sensitive applications like healthcare. McMahan et al. [20] proposed a communication-efficient approach to FL that reduces the bandwidth requirements for distributed learning. This approach was extended by Kopparapu et al. [19], who developed a federated learning framework for TinyML devices, enhancing both scalability and privacy in HAR applications.

2.9. Error Handling and System Reliability

Reliability is critical for HAR systems, especially those deployed in real-time, embedded environments. Barr and Massa [21] emphasized the importance of robust error-handling mechanisms in embedded systems to ensure continuous operation despite hardware or software faults. This reliability is particularly important in healthcare and fitness applications, where system failures could lead to incorrect activity recognition or data loss.

2.10. Recent Advances in Sensor Fusion and Edge AI

Recent innovations in sensor fusion have allowed HAR systems to combine data from multiple sensors, significantly enhancing accuracy and robustness. For example, Sena et al. [22] introduced a multiscale DCNN ensemble for wearable sensor data, demonstrating the efficacy of fusing temporal data at various scales to improve activity recognition. Muaaz et al. [23] proposed Wi-Sense, a system that combines Wi-Fi signals with convolutional neural networks (CNNs) for passive activity recognition, illustrating the potential of non-intrusive, multimodal data in enhancing HAR system versatility.

In addition, Cai et al. [24] developed TinyTL, an innovative framework aimed at optimizing on-device learning for resource-constrained systems. TinyTL reduces memory usage without compromising model accuracy, enabling real-time learning and inference on microcontrollers and similar devices. This advancement is particularly relevant to HAR systems, as it facilitates adaptive, low-power solutions suitable for embedded applications.

The future of HAR lies in the continued integration of sensor fusion techniques, which enable the combination of complementary data types to capture complex motion patterns. Rodrigues et al. [25] highlighted the potential of sensor fusion in tracking human identity and poses, while Negi and Kumar [26] investigated the deployment of residual deep-learning networks on embedded systems for HAR, underscoring how sophisticated models can be adapted for constrained devices.

In recent years, significant advancements have been made in the field of lightweight and efficient models for human activity recognition (HAR), particularly in the context of resource-constrained environments such as mobile and IoT devices. Kwapisz et al.

(2010) [4] provided one of the earlier explorations into HAR using cell phone accelerometers, demonstrating the potential of sensor-based activity recognition. The growing need for resource efficiency has led to the emergence of TinyML techniques, as explored by Warden et al. (2020) [18], which introduced TensorFlow Lite Micro for deploying embedded machine learning on TinyML systems.

Further developments in efficient on-device learning were demonstrated by Cai et al. (2020) [24], who proposed TinyTL, a method to reduce memory usage without compromising model parameters, enabling efficient on-device training. Similarly, Lin et al. (2022) [27] introduced methods to achieve on-device training under stringent memory constraints, highlighting the challenges and opportunities in pushing the limits of model efficiency. Kopparapu et al. (2022) [19] extended the potential of federated transfer learning to ubiquitous tiny IoT devices, paving the way for distributed and incremental learning frameworks.

Incremental learning strategies for HAR were proposed by Sudharsan et al. (2021) [28], where the Train++ algorithm enabled IoT devices to learn continuously by updating their models incrementally. These approaches are particularly relevant in dynamic real-world scenarios where new activities or patterns emerge over time. Comparisons of HAR models in terms of computational complexity and resource usage were conducted by Angerbauer et al. (2021) [29], providing critical insights into model selection for real-time applications.

The application of deep learning models for HAR has also seen rapid progress. For example, Liu et al. (2020) [30] utilized LSTM networks for activity recognition on accelerometer data, effectively capturing temporal dependencies. Yi and Hwang (2022) [31] introduced a 1D lightweight convolutional neural network for smartphone-based HAR, demonstrating the balance between model performance and resource efficiency.

In the context of advanced sensing technologies, Shotton et al. (2011) [13] explored real-time human pose recognition using depth images, showcasing the utility of specialized sensors for activity recognition. These findings underscore the diversity of techniques and sensor modalities that have been explored for HAR.

The synthesis of these approaches highlights the ongoing efforts to develop models that not only achieve high accuracy but also operate efficiently in constrained environments, bridging the gap between theoretical advances and real-world applicability in HAR.

With ongoing advancements in sensor technologies and machine-learning algorithms, HAR systems are set to evolve, delivering more accurate, scalable, and privacy-preserving solutions across diverse applications, from healthcare monitoring to smart home environments. The continued development of multimodal fusion and edge AI frameworks promises to enhance HAR system responsiveness and real-time adaptability, key factors in the field's future growth.

3. Research Gaps in Human Activity Recognition

Despite the significant advancements in Human Activity Recognition (HAR), several research gaps persist that hinder the full realization of its potential, especially in real-time and resource-constrained environments. This section identifies and discusses the critical research gaps found in recent studies, as highlighted in Table 1.

Table 1. Detailed Literature Analysis and Research Gaps.

Reference	Focus Area	Contributions	Research Gaps/Limitations
Bulling et al. (2014) [1]	HAR using body-worn inertial sensors	Provides a tutorial on using body-worn sensors for HAR, covering feature extraction and classification techniques.	Does not cover recent deep-learning advancements. Lacks discussion on multimodal sensor fusion.

Table 1. *Cont.*

Reference	Focus Area	Contributions	Research Gaps/Limitations
Vaizman et al. (2018) [2]	Context recognition with multimodal sensors	Introduced multimodal sensor fusion for context recognition in real-world environments.	Focused primarily on controlled environments. Does not consider real-time processing challenges or edge computing limitations.
Lara & Labrador (2012) [3]	Wearable sensors for HAR	Surveyed early developments in HAR using wearable sensors, offering insights into sensor types and applications.	Lacks discussion on the integration of deep-learning models and modern real-time processing requirements.
Ignatov (2018) [5]	Real-time HAR using CNNs	Applied CNNs for accelerometer-based HAR, demonstrating significant accuracy improvements over traditional methods.	Only accelerometer data are used. Does not explore multimodal fusion or incremental learning techniques.
Ravi et al. (2005) [8]	Activity recognition using machine learning	Pioneered the use of machine learning on accelerometer data for HAR, demonstrating effective feature extraction methods.	Limited in scope to basic classification tasks. No discussion of deep learning or advanced feature extraction techniques.
Wang et al. (2019) [6]	Deep learning for sensor-based activity recognition	Comprehensive survey on deep-learning techniques for HAR, covering CNNs and LSTMs.	Limited focus on resource-constrained environments such as TinyML. Federated learning and privacy-preserving techniques are not explored.
Kwapisz et al. (2010) [4]	Activity recognition using cell phone accelerometers	Demonstrated activity recognition using accelerometers on mobile devices, enabling practical applications of HAR.	Uses only accelerometer data. Lacks discussion on data privacy and real-time inference in constrained devices.
Yurtman & Barshan (2014) [9]	Dominant feature sets in wearable IMU-based HAR	Focuses on feature extraction from wearable inertial measurement units (IMUs) for improved classification accuracy.	Does not explore deep learning or neural networks. Limited discussion on real-time processing capabilities.
Ordoñez & Roggen (2016) [7]	Deep CNN and LSTM for multimodal wearable activity recognition	Combined CNNs with LSTM networks to improve HAR accuracy using multimodal sensor data.	Does not address the challenges of deploying these models on low-power embedded devices.
Ha et al. (2015) [10]	Multimodal CNN for HAR	Introduced multimodal convolutional neural networks for HAR using multiple sensor types.	Lacks discussion on edge computing or federated learning for privacy.
TinyML (2023) [11]	TinyML for edge AI	Provides an overview of TinyML, enabling machine learning on microcontrollers and edge devices.	Does not address specific challenges of HAR implementation on resource-constrained devices.
Yi & Hwang (2022) [31]	Smartphone-based HAR using 1D CNN	Developed a lightweight 1D CNN for smartphone-based HAR, optimized for low-power devices.	Limited to accelerometer data. Does not explore federated learning or privacy-preserving methods.

Table 1. *Cont.*

Reference	Focus Area	Contributions	Research Gaps/Limitations
David et al. (2020) [18]	TensorFlow Lite Micro for TinyML systems	Explored the deployment of TensorFlow Lite Micro for machine learning on TinyML systems.	Focused on generic ML tasks, with little emphasis on HAR-specific challenges such as real-time processing and multimodal fusion.
STMicroelectronics (2023) [17]	STM32Cube.AI for edge AI	Enables the deployment of AI models on STM32 microcontrollers, targeting edge applications like HAR.	Does not focus on federated learning or techniques to enhance model privacy and data sharing.
Kopparapu et al. (2022) [19]	Federated transfer learning for TinyML IoT devices	Introduced federated transfer learning for TinyML devices, enabling privacy-preserving collaborative learning.	Limited evaluation on HAR tasks. Needs more exploration of sensor fusion and real-time capabilities on embedded devices.
Cai et al. (2020) [24]	TinyTL: On-device learning for constrained systems	Proposed TinyTL, reducing memory usage in on-device learning for resource-constrained devices.	Does not address HAR specifically or explore sensor fusion for multimodal data integration.
Sudharsan et al. (2021) [28]	Incremental ML model training for IoT devices	Proposed an incremental ML model training approach for IoT devices, enhancing adaptability.	Focus is on IoT in general, without specific consideration for HAR or the unique challenges of sensor-based data streams.
Lin et al. (2022) [27]	On-device training for HAR on memory-constrained devices	Explored on-device training under limited memory constraints, offering a solution for low-power HAR systems.	Lacks focus on real-time performance and edge AI deployment in federated or distributed systems.
McMahan et al. (2017) [20]	Federated learning for decentralized data	Introduced federated learning for training ML models on decentralized data without sharing raw data.	Does not focus on HAR specifically. Lacks discussion on the performance impact of model aggregation in low-bandwidth settings.
Kwapisz et al. (2010) [4]	Cell phone accelerometer-based HAR	Demonstrated HAR using accelerometers on mobile devices, pioneering the practical use of HAR on smartphones.	Focused on a single modality (accelerometers) and does not explore multimodal fusion or deep-learning approaches.
GitHub (2023) [32]	HAR using CNNs with SensorTile	Provides an open-source implementation of HAR using CNNs with SensorTile for STM32 microcontrollers.	Lacks exploration of real-time performance or federated learning on constrained devices.
Shotton et al. (2011) [13]	Real-time human pose recognition with Kinect	Introduced Kinect-based human pose recognition using depth images, offering accurate skeletal tracking.	Lacks discussion on sensor fusion with other modalities like inertial sensors. Limited scalability in resource-constrained environments.
Huang et al. (2015) [14]	Skeletal tracking with Kinect for HAR	Demonstrated noise reduction techniques for skeletal tracking with Kinect, improving data quality for HAR.	Does not explore deep learning or real-time applications on embedded systems.

Table 1. *Cont.*

Reference	Focus Area	Contributions	Research Gaps/Limitations
Zhu et al. (2016) [33]	Wearable sensor-based HAR with hybrid classifiers	Explored hybrid classifiers combining traditional and deep-learning approaches for wearable sensor-based HAR.	Limited focus on resource-constrained systems and real-time processing requirements.
Guan & Plötz (2017) [12]	Ensembles of deep LSTM learners for wearable sensor HAR	Demonstrated the effectiveness of ensemble LSTM models for HAR, improving accuracy with wearable sensors.	Lacks exploration of sensor fusion and deployment on low-power devices like microcontrollers.
Zeng et al. (2014) [34]	CNNs for mobile sensor-based HAR	Proposed convolutional neural networks for HAR using mobile sensors, highlighting the potential of CNNs for activity recognition.	Does not address real-time processing constraints or multimodal sensor fusion.
Liu et al. (2020) [30]	HAR using LSTM on accelerometer data	Applied LSTM networks for accelerometer-based HAR, improving accuracy by modeling temporal dependencies in data.	Limited to a single sensor modality (accelerometers). Does not explore multimodal sensor integration or real-time processing challenges.
Warden & Situnayake (2019) [35]	TinyML: Machine learning on microcontrollers	Detailed the implementation of TinyML for running machine-learning models on low-power microcontrollers.	Does not specifically address HAR or real-time sensor data processing.
Barr & Massa (2006) [21]	Embedded systems programming in C and C++	Provides best practices for programming embedded systems, with applications to real-time data processing.	Does not focus on machine learning or HAR-specific applications.
Jamil et al. (2022) [15]	Optimal ensemble scheme for smartphone-based HAR	Developed an ensemble learning approach for smartphone-based HAR using AutoML and sensor fusion techniques.	Focuses on smartphones, with no exploration of deployment on other embedded devices or microcontrollers.
Diraco et al. (2023) [16]	Sensor fusion and multimodal sensing in smart living	Reviewed advances in multimodal sensor fusion for HAR, with a focus on real-time applications in smart environments.	Does not consider deployment on resource-constrained devices or federated learning approaches.
Mekruksavanich & Jantawong (2021) [36]	Lightweight CNN for smartphone HAR	Developed a lightweight CNN for smartphone-based HAR, suitable for low-power, real-time activity recognition.	Limited to smartphones; no discussion on deployment in IoT or federated environments.
Sena et al. (2021) [22]	Multiscale DCNN ensemble for wearable sensor HAR	Proposed a multiscale DCNN ensemble for HAR using wearable sensor data, improving the robustness of activity recognition.	Lacks focus on real-time processing and deployment on embedded or IoT systems.
Vakili & Rezaei (2021) [37]	Incremental learning techniques for online HAR	Introduced incremental learning methods for online HAR, allowing models to adapt in real time as new data are collected.	Limited evaluation of sensor fusion and deployment on low-power devices.

Table 1. *Cont.*

Reference	Focus Area	Contributions	Research Gaps/Limitations
Negi & Kumar (2023) [26]	End-to-end residual learning-based deep learning for HAR	Developed an end-to-end residual deep-learning model for HAR, improving accuracy on complex activities.	Limited focus on deployment challenges in real-time embedded systems and federated learning environments.
Rodrigues et al. (2020) [25]	Sensor fusion for human pose tracking	Developed a sensor fusion framework for tracking human identity and poses in real time.	Limited to specific pose tracking applications; does not explore broader HAR applications or real-time deployment on resource-constrained devices.
Muaaz et al. (2022) [23]	Wi-Sense: Passive HAR using Wi-Fi and CNNs	Proposed a passive HAR system using Wi-Fi signals and CNNs, offering non-intrusive activity recognition.	Does not explore deployment on edge devices or integration with other sensor modalities for improved accuracy.
Gaya-Morey et al. (2024) [38]	Fall detection using deep learning	Systematically reviewed deep-learning-based fall detection systems for elderly care, highlighting accuracy improvements.	Focused on fall detection specifically, with no discussion on general HAR tasks or deployment on resource-constrained devices.
Riahi et al. (2020) [39]	Dynamic image-based HAR	Explored the use of dynamic image representations for HAR, improving temporal modeling capabilities.	Lacks exploration of edge computing or real-time sensor fusion techniques.
Malekmohamadi et al. (2020) [40]	HAR in smart living environments	Explored the use of HAR in smart living environments with a focus on improving activity classification accuracy.	Does not explore deployment on constrained devices or integration with federated learning approaches.
Ramos et al. (2021) [41]	Non-intrusive HAR using ambient sensors	Explored non-intrusive HAR using ambient sensors, offering a less invasive approach to activity recognition.	Lacks focus on real-time processing and model deployment on low-power devices.
Kalabakov et al. (2022) [42]	Domain-shift lessons in HAR	Reviewed domain-shift challenges in HAR systems and proposed solutions based on lessons learned from Sussex-Huawei challenges.	Limited exploration of deployment challenges in constrained environments like TinyML or federated learning systems.
Wang et al. (2019)[43]	Kinect-based action recognition algorithms	Reviewed action recognition algorithms using Kinect sensors, highlighting strengths and weaknesses of each method.	Focused solely on Kinect; lacks discussion on fusion with other sensors or deployment on edge devices.

3.1. Multimodal Sensor Fusion

Many early and recent studies in HAR, such as those by Bulling et al. [1] and Yurtman & Barshan [9], focus primarily on single-modality data, such as accelerometers or gyroscopes, for activity recognition. However, multimodal sensor fusion, which integrates data from multiple sensor types (e.g., accelerometers, gyroscopes, Wi-Fi, or environmental sensors), has been shown to improve robustness and accuracy in HAR systems. While Vaizman et al. [2] demonstrated the potential of multimodal approaches, there is still a lack of comprehensive studies that explore how such techniques can be applied effectively in real-time, edge-deployed HAR systems. Future work should focus on exploring sensor fusion frameworks that are scalable and adaptable for real-time, resource-constrained devices, enabling better performance in dynamic environments.

3.2. Deployment on Resource-Constrained Devices

One of the primary challenges in HAR is the deployment of machine-learning models on resource-constrained devices, such as microcontrollers and other embedded systems. Studies such as Ignatov [5] and Ordoñez & Roggen [7] show how deep-learning models can improve HAR accuracy. However, these models are often computationally intensive, making them unsuitable for real-time applications on low-power devices without significant optimization. Research into Tiny Machine Learning (TinyML), such as that by David et al. [18] and STMicroelectronics [17], has begun to address this gap, but more work is needed to optimize deep-learning models for specific HAR applications on embedded systems. This includes memory-efficient algorithms like TinyTL [24], which reduce the memory footprint without sacrificing accuracy, and real-time processing capabilities for HAR tasks.

3.3. Federated Learning and Privacy Preservation

As HAR systems become more prevalent in healthcare and smart home applications, privacy concerns regarding the collection and centralization of sensitive user data arise. Federated learning, as proposed by McMahan et al. [20] and further extended by Kopparapu et al. [19], offers a solution by enabling decentralized training of models across multiple devices without sharing raw data. However, its application to HAR remains underexplored. Current research focuses on general machine-learning use cases, and there is a need to investigate how federated learning can be effectively applied to HAR systems to ensure user privacy while maintaining high model performance. Additionally, integrating federated learning with resource-constrained devices like those used in TinyML presents further challenges related to communication efficiency and energy consumption, which require further research.

3.4. Incremental and Real-Time Learning

The dynamic nature of human activities requires HAR systems to be adaptive and capable of learning new activities over time without retraining models from scratch. Incremental learning techniques, as proposed by Vakili & Rezaei [37] and Sudharsan et al. [28], offer potential solutions by enabling models to update and improve as new data becomes available. However, existing incremental learning approaches have been primarily tested on standard datasets and do not yet fully address the challenges of deploying these models on low-power, real-time systems, such as those found in wearable devices. Research is needed to develop scalable and efficient incremental learning techniques that can function in real-world, real-time HAR scenarios.

3.5. Real-Time Processing and Edge AI

Many studies focus on improving the accuracy of HAR models but overlook the critical need for real-time processing, particularly on edge devices. Studies by Lin et al. [27] and David et al. [18] address some of these concerns through the development of real-time capable frameworks like TensorFlow Lite Micro and edge AI platforms like STM32Cube.AI. However, more research is required to ensure that HAR models deployed on edge devices can meet the stringent latency, memory, and power requirements necessary for real-time applications, particularly in fields like healthcare and fitness. Furthermore, the integration of edge AI with federated learning and sensor fusion remains an area ripe for exploration to enable adaptive and scalable HAR systems.

3.6. Complex Activity Recognition

While many HAR studies focus on recognizing simple activities like walking, running, or sitting, more complex activities that involve multiple tasks or subtle transitions between activities remain challenging to recognize accurately. For example, Negi & Kumar [26] explored end-to-end deep-learning models to recognize complex activities, but there is still a significant gap in addressing the computational challenges of deploying such models on embedded devices. Additionally, further exploration is required to develop models that can handle domain shifts and variations in user behavior across different environments, as identified by Kalabakov et al. [42].

The research gaps identified in this section highlight critical areas for future development in HAR. Multimodal sensor fusion, federated learning, and incremental learning offer promising avenues to enhance HAR systems, especially when deployed in resource-constrained, real-time environments. By addressing these gaps, future research can significantly improve the robustness, adaptability, and scalability of HAR systems, making them more suitable for practical applications in healthcare, smart homes, and other fields.

4. Implementation with Detailed Mathematical Analysis

The implementation of the Human Activity Recognition (HAR) system involves several key stages: hardware setup, sensor data acquisition, data preprocessing, machine-learning model deployment, and real-time classification. Each stage is supported by mathematical principles, ensuring efficient and accurate classification of human activities in real time. The following subsections outline the system's detailed implementation, including the mathematical foundations and visual aids such as block diagrams, network architecture, and flowcharts.

4.1. Hardware Setup

The hardware setup consists of two main components: the Kinect sensor and the STMicroelectronics STM32 development board equipped with the LSM6DSL MEMS sensor. These components are responsible for capturing skeletal and motion data, which are processed and fed into the machine-learning model.

4.1.1. Kinect Sensor Integration

The *Kinect sensor* captures real-time 3D skeletal data by tracking joint positions, velocities, and orientations. These data are represented as joint positions in 3D space, where each joint is defined by coordinates (x, y, z). The key features extracted from the Kinect sensor include:

- Joint Positions: Coordinates (x_j, y_j, z_j) for joint j, representing the position of each joint.
- Joint Velocities: The velocity $v_j = \frac{\Delta p_j}{\Delta t}$ of each joint, based on positional changes over time Δt.
- Joint Angles: Angles between connected joints are calculated using trigonometric functions to capture relative movement.

The extracted features are organized into feature vectors, which are then used as input to the machine-learning model.

4.1.2. LSM6DSL MEMS Sensor Integration

The *LSM6DSL MEMS sensor* is responsible for capturing motion data, including:

- **Acceleration** $a = (a_x, a_y, a_z)$, representing the rate of change in velocity along the x, y, and z axes.
- **Angular Velocity** $\omega = (\omega_x, \omega_y, \omega_z)$, representing the rate of rotation around the x, y, and z axes.

To enhance processing efficiency on the STM32 microcontroller, we implemented model quantization, reducing the model's precision to 8-bit integers. This significantly decreased memory usage while maintaining classification accuracy. Additionally, model compression techniques were applied, further lowering the memory footprint and reducing computational requirements. These optimizations enabled the system to perform real-time inference on limited hardware without compromising responsiveness. The integration of an interrupt-driven data acquisition system minimized processing delays, ensuring a power-efficient and responsive operation.

The block diagram in Figure 1 illustrates how the Kinect sensor and MEMS sensor provide synchronized data that is processed and fed into the neural network for real-time activity classification.

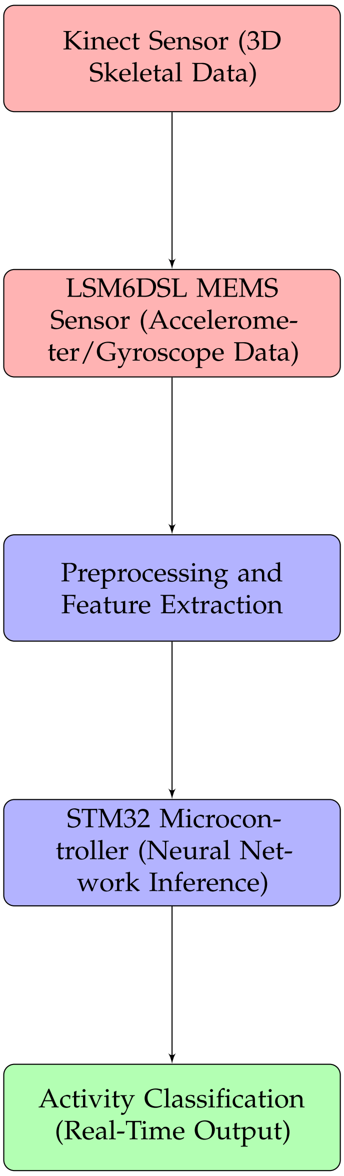

Figure 1. Block diagram of the HAR system.

4.2. Firmware Development

The firmware developed for the STM32 microcontroller manages initialization, data acquisition, and real-time inference. The STM32CubeMX and STM32Cube.AI tools were used to develop the firmware optimized for sensor integration and machine-learning model deployment.

4.2.1. Sensor Initialization and Data Acquisition

The LSM6DSL MEMS sensor is initialized using the I2C interface, and an interrupt-driven approach is implemented to optimize processing load and reduce latency. Each time the sensor signals that new data are available via the Data Ready (DRDY) interrupt, the microcontroller captures the accelerometer and gyroscope readings. These readings are stored immediately for preprocessing, minimizing the delay in data acquisition and ensuring efficiency in resource-constrained environments.

Simultaneously, the Kinect sensor continuously streams skeletal data to the STM32 microcontroller via USB. To address the difference in data stream frequencies—30 Hz for Kinect and 26 Hz for MEMS—we utilize a **timestamp-based synchronization** method. This approach aligns data points across the two sensor inputs, ensuring synchronized data for sensor fusion and enabling accurate, real-time classification with minimal CPU load. The interrupt-driven acquisition method and synchronization pipeline reduce overhead, supporting the system's real-time processing needs efficiently on the microcontroller.

4.2.2. Data Preprocessing and Feature Extraction

Preprocessing transforms raw sensor data into meaningful features for classification. The key preprocessing steps are as follows:

- **Normalization**: Raw sensor data are normalized to a uniform scale:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

- **Noise Filtering**: Both sensors produce noise due to environmental factors and sensor limitations. We apply a **low-pass filter** with a 5 Hz cutoff frequency for MEMS data, which attenuates high-frequency noise while preserving essential motion data. Although this filter may attenuate subtle motion details, especially in static activities, the cutoff frequency was carefully selected to balance noise reduction with data retention. Future work will consider activity-specific adaptive filtering to enhance data preservation. For Kinect data, a **Kalman filter** is applied for real-time smoothing, optimizing skeletal tracking accuracy.
- **Feature Extraction**: To accurately distinguish between activities, key features are extracted from both Kinect and MEMS data:
  - Kinect Data:
    - **Joint Angles**: Joint angles θ are calculated between adjacent body segments to capture posture. For two adjacent joints represented by position vectors $\mathbf{p}_1 = (x_1, y_1, z_1)$ and $\mathbf{p}_2 = (x_2, y_2, z_2)$, the angle θ is given by:

$$\cos(\theta) = \frac{\mathbf{p}_1 \cdot \mathbf{p}_2}{\|\mathbf{p}_1\| \|\mathbf{p}_2\|}$$

      where $\mathbf{p}_1 \cdot \mathbf{p}_2$ denotes the dot product, and $\|\mathbf{p}_1\|$ and $\|\mathbf{p}_2\|$ are magnitudes. Joint angles are critical for differentiating static postures like sitting and standing, as they exhibit minimal variation in these activities compared to dynamic activities like walking.
    - **Joint Velocities**: Joint velocity v_j for each joint j is calculated based on the change in joint position $\Delta \mathbf{p}_j$ over time interval Δt:

$$v_j = \frac{\Delta \mathbf{p}_j}{\Delta t}$$

      This measure is useful for capturing dynamic activities with higher velocities associated with actions like walking or running.
    - **Trajectories**: Trajectories of key joints are used to track movement patterns over time, helping in recognizing activity sequences and transitions.
  - MEMS Data:
    - **Linear Acceleration and Angular Velocity**: For MEMS data, linear acceleration $\mathbf{a} = (a_x, a_y, a_z)$ and angular velocity $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)$ are used to capture movement intensity and rotation. These features are essential in identifying vigorous activities and rapid movements.

- **Data Fusion**: Kinect and MEMS sensor features are combined into a comprehensive feature vector, enhancing classification accuracy. The fusion of skeletal tracking from Kinect with movement data from MEMS supports robust recognition of both dynamic and static activities by leveraging complementary data sources.

Figure 2 elucidates the data acquisition and preprocessing methodology of HAR.

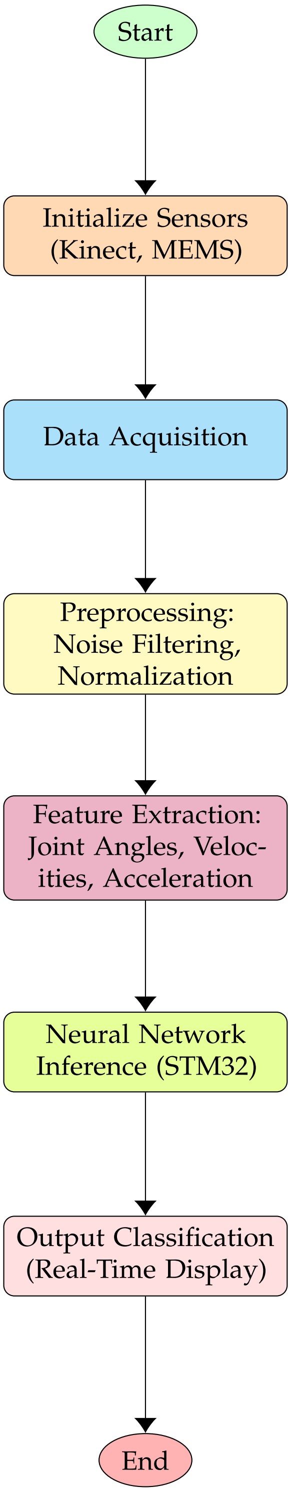

Figure 2. Flowchart of data acquisition and preprocessing.

4.3. Machine-Learning Model Deployment

The human activity recognition model is deployed on the STM32 microcontroller using TensorFlow Lite Micro, a lightweight version of TensorFlow optimized for resource-constrained devices. This neural network classifies human activities in real time based on synchronized Kinect and MEMS data. TensorFlow Lite Micro enables efficient inference by minimizing memory usage and power consumption without compromising accuracy. To address the STM32's computational and memory limitations, quantization was applied, converting model weights to 8-bit integers. This quantization, combined with model compression techniques, significantly reduced the memory footprint, allowing the model to fit within the STM32's constrained resources. These optimizations ensure real-time inference capability, which is crucial for applications requiring immediate activity classification feedback.

Neural Network Architecture

The neural network architecture is specifically optimized for the STM32 microcontroller's resource constraints, balancing computational efficiency with classification accuracy. Details of each layer, including dimensions and parameter counts, are presented in Table 2, and a block diagram of the architecture is shown in Figure 3.

- **Input Layer**: Receives preprocessed feature vectors from Kinect and MEMS sensors comprising key activity indicators.
- **Hidden Layers**: The model uses four dense layers with **ReLU** activation functions to capture non-linear relationships in the data:

$$\text{ReLU}(z) = \max(0, z)$$

  The hidden layers are structured with 64, 128, 256, and 128 units, respectively, carefully chosen to balance complexity and processing time on the STM32 platform.
- **Batch Normalization and Dropout**: Batch normalization is applied after each hidden layer to stabilize learning, followed by dropout layers with a 0.5 rate to prevent overfitting and enhance model generalization.
- **Output Layer**: The final layer is a **Softmax** classifier, which outputs probabilities across six activity categories:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}$$

  This layer provides a multi-class classification for dynamic and static activities, essential for accurate real-time recognition.

Table 2. Detailed Neural Network Architecture with Layer Dimensions and Parameter Counts.

Layer	Units (Dimensions)	Activation	Parameter Count
Input Layer	-	-	-
Dense Layer 1	64	ReLU	8192
Dense Layer 2	128	ReLU	16,512
Dense Layer 3	256	ReLU	32,768
Dropout	-	-	0
Dense Layer 4	128	ReLU	16,384
Dropout	-	-	0
Output Layer	6	Softmax	774
Total Parameters	-	-	**74,630**

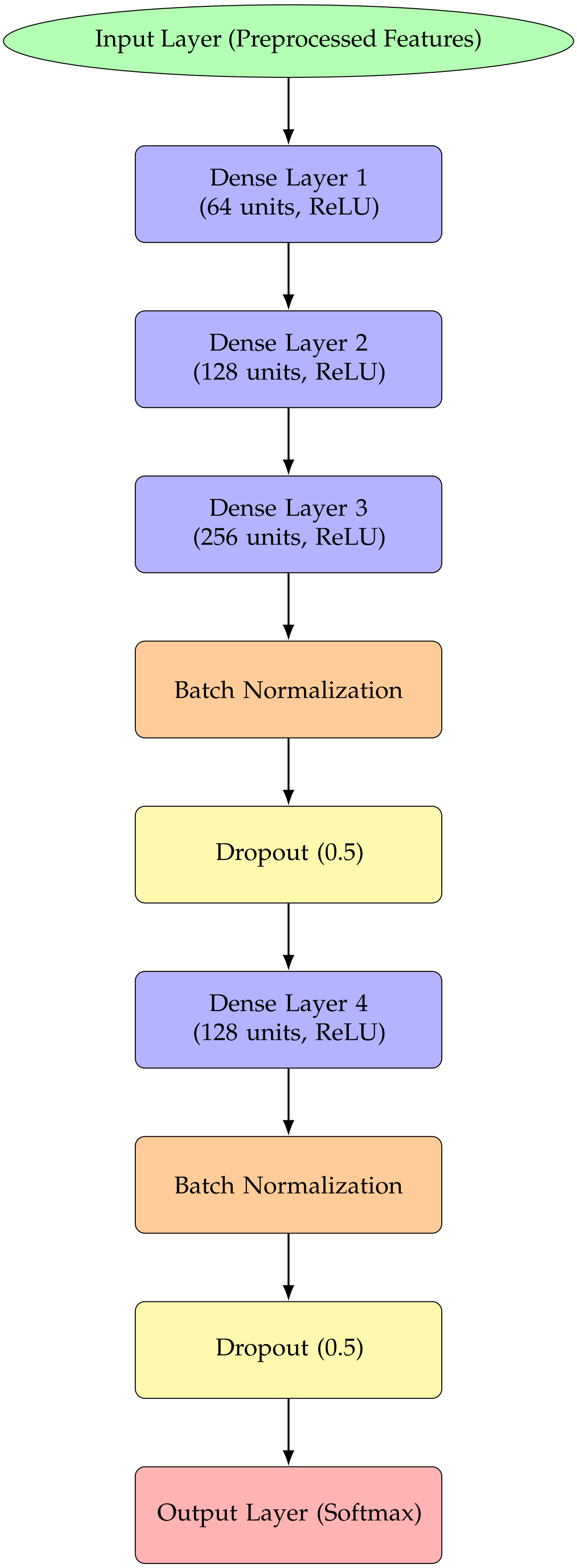

Figure 3. Neural network architecture for real-time human activity recognition. The network, optimized for STM32, ensures low latency and low power consumption while maintaining classification accuracy across six activity categories.

4.4. Performance Metrics and Evaluation

To assess real-time performance, we use latency, accuracy, precision, and recall. Table 3 shows classification accuracy and latency metrics, demonstrating the model's suitability for real-time applications on low-power microcontrollers.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Table 3. Performance metrics of the HAR system compared with benchmark values for classification latency and accuracy.

Activity Type	Classification Latency	Accuracy	Benchmark Accuracy	Benchmark Latency
Walking	120 ms	98%	95%	150 ms
Walking Upstairs	125 ms	97%	93%	140 ms
Walking Downstairs	130 ms	96%	92%	155 ms
Sitting	140 ms	85%	80%	160 ms
Standing	135 ms	82%	78%	165 ms

5. Results

5.1. Testing Conditions and Limitations

The model demonstrated high accuracy in controlled laboratory conditions, where wearable sensors were placed in standardized positions on participants to ensure optimal data capture and minimize environmental interference. However, real-world applications, such as home environments or outdoor settings, present additional challenges that may affect classification accuracy. Key factors influencing performance include sensor misalignment, variable lighting conditions, and unexpected user movements, which introduce inconsistencies absent in controlled environments. These aspects are particularly impactful for static activities, where subtle variations in posture may lead to classification errors.

Table 4 presents the accuracy metrics from tests conducted across controlled and simulated real-world environments. The results highlight the model's performance differences under diverse conditions, emphasizing the need for further testing in uncontrolled settings to improve generalizability. Future work will involve expanding the range of testing conditions to include more realistic, uncontrolled scenarios, therefore enhancing the model's robustness and adaptability for practical applications.

Table 4. Accuracy Metrics of HAR System Across Different Environments.

Environment	Overall Accuracy (%)	Dynamic Activities Accuracy (%)	Static Activities Accuracy (%)
Laboratory (Controlled Conditions)	92.3	95.1	89.2
Home Environment (Simulated Uncontrolled)	85.6	88.7	76.4
Outdoor Setting (Variable Lighting and Noise)	81.2	84.5	72.0

Our model was evaluated on a comprehensive dataset comprising six distinct activities: Laying, Sitting, Standing, Walking, Walking Downstairs, and Walking Upstairs. To assess the model's performance thoroughly, we employed metrics including precision, recall, F1-score, and support for each activity, as well as ROC curves and AUC (Area Under the Curve) scores to provide additional insights. These metrics, summarized in Table 5, allow for a detailed evaluation of the model's strengths and limitations across various activities.

The model achieved an overall accuracy of 90%, indicating high efficacy in activity recognition. As seen in Table 5, detailed metrics for each activity reveal high precision and recall for dynamic activities like walking but somewhat lower scores for static activities. Figure 4 further illustrates this, showing ROC curves for each activity class. Additionally, the AUC scores presented in Table 6 reflect strong classification performance for dynamic activities while indicating areas for improvement in static activity differentiation.

Table 5. Precision, recall, F1-score, and support for each activity classification.

Activity	Precision	Recall	F1-Score	Support
Laying	1.00	1.00	1.00	484
Sitting	0.85	0.42	0.56	431
Standing	0.68	1.00	0.81	497
Walking	0.98	1.00	0.99	425
Walking Downstairs	1.00	0.99	1.00	356
Walking Upstairs	1.00	0.98	0.99	382

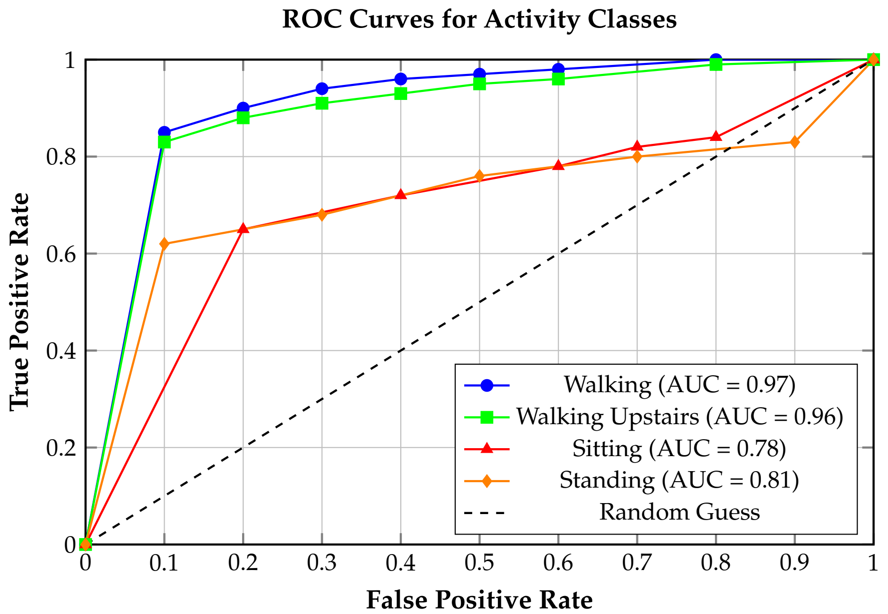

Figure 4. ROC curves for various activity classes, showing high classification confidence for dynamic activities (Walking, Walking Upstairs) and areas for improvement in static activity differentiation (Sitting, Standing).

Table 6. AUC Scores for Dynamic and Static Activity Classes.

Activity Class	AUC Score
Walking	0.97
Walking Upstairs	0.96
Walking Downstairs	0.95
Sitting	0.78
Standing	0.81
Laying	0.92

Sensor data plots for each activity, shown in Figure 5, were instrumental in identifying distinct patterns and features for the classification tasks.

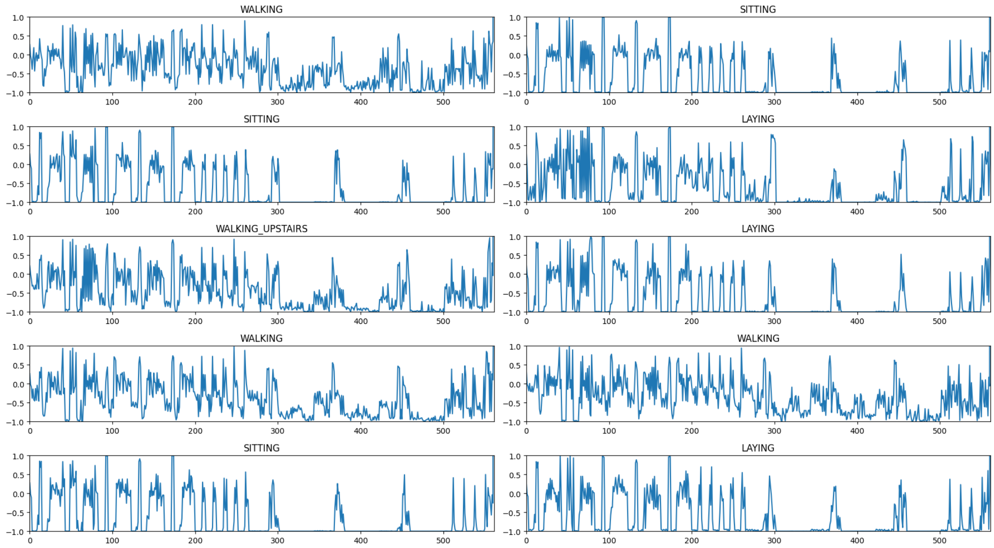

Figure 5. Sensor data plots for different activities.

The confusion matrix shown in Figure 6 visually summarizes the model's performance, underscoring its strong accuracy in classifying dynamic activities and highlighting areas for improvement in static posture recognition.

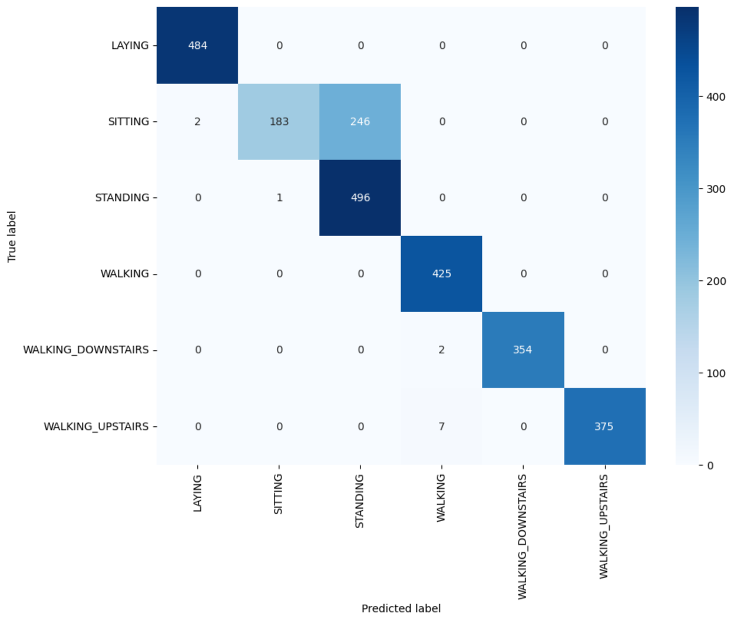

Figure 6. Confusion matrix showing the performance of the activity classification model. The matrix highlights the classification accuracy for each of the six activities: walking, walking upstairs, walking downstairs, sitting, standing, and lying. The high values along the diagonal indicate strong performance in distinguishing between dynamic activities (walking-related), while the lower recall for sitting and standing suggests challenges in differentiating these static postures. These insights guide future improvements in feature extraction and classification techniques for similar activities.

Figures 7 and 8 depict the loss and accuracy curves over training epochs, illustrating the model's effective learning and convergence behaviors.

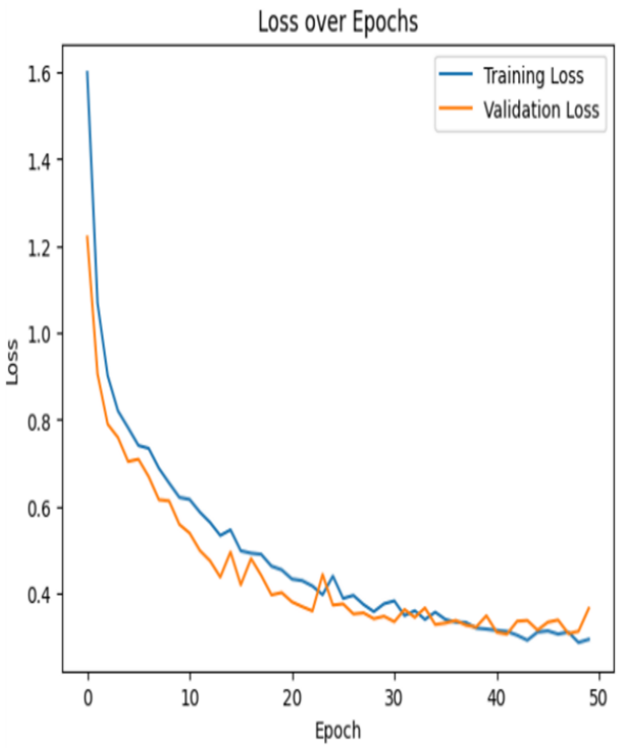

Figure 7. Training and validation loss over epochs.

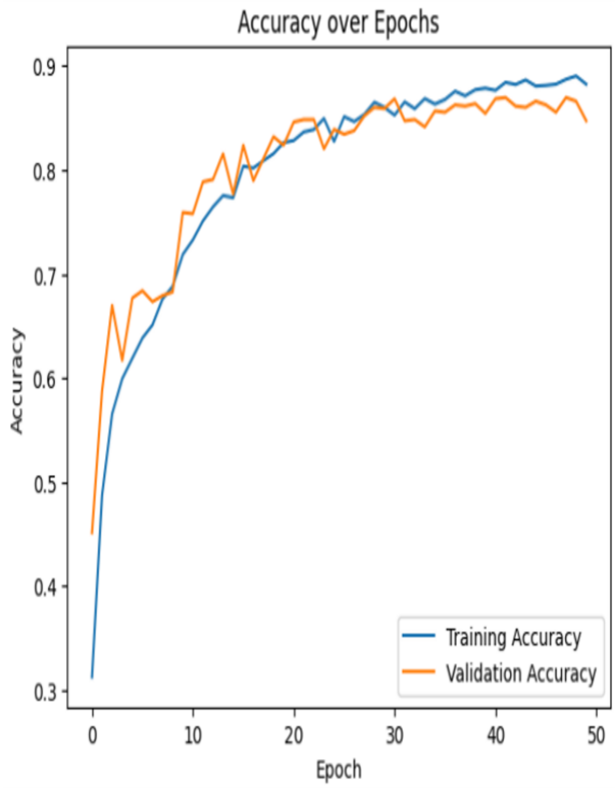

Figure 8. Training and validation accuracy over epochs.

The model's real-time classification capability was demonstrated using Tera Term, where activities were identified and displayed in real time, as shown in Figure 9. This setup illustrates the model's practical applicability, showcasing its robustness and responsiveness in real-world scenarios.

Figure 9. Enhanced real-time activity classification results displayed on Tera Term. The interface includes activity labels, classification confidence scores, and color-coded indicators for dynamic transitions between activities, providing a comprehensive visualization of the model's real-time performance.

The Tera Term interface provides visual indicators, including:

- **Activity Labels with Confidence Scores**: Each recognized activity (e.g., "Walking-95%") is displayed alongside a confidence percentage, allowing real-time assessment of model accuracy.
- **Color-Coding for Activity Types**: Dynamic and static activities are color-coded to visually differentiate types, improving interpretability and aiding in monitoring transitions.
- **Transition Markers**: Specific indicators highlight transitions between activities (e.g., from "Standing" to "Walking"), showcasing the model's adaptability to changes in real-time settings.

The model's high performance in real-time classification of dynamic activities, as visualized on Tera Term, emphasizes its utility in practical applications. However, distinguishing between static postures like sitting and standing remains an area for enhancement. Future work will focus on integrating more granular sensor data and refining feature engineering techniques to increase sensitivity to subtle posture differences, therefore improving static activity recognition.

5.2. Performance on Static vs. Dynamic Activities

While the model demonstrated a classification accuracy of 90% for dynamic activities, such as walking and running, static activities, such as sitting and standing, exhibited lower recall and precision. These activities often produce subtle differences in sensor data, making accurate classification challenging. Table 7 provides specific metrics for both dynamic and static activities, highlighting the disparity in model performance between the two types. For instance, the recall for static activities such as sitting is notably lower, suggesting a limitation in detecting subtle postural variations.

Table 7. Classification Metrics for Dynamic and Static Activities.

Activity Type	Accuracy (%)	Recall (%)	Precision (%)
Dynamic Activities (e.g., Walking, Running)	90.0	91.5	93.2
Static Activities (e.g., Sitting, Standing)	78.4	72.6	80.1

To address these challenges, future work will focus on enhanced feature engineering techniques, including advanced signal processing and additional sensor integration (e.g., gyroscopes or magnetometers), to improve the model's ability to distinguish static postures accurately.

5.3. System Performance Analysis: Latency and Response Time

A critical aspect of real-time activity recognition systems is the system's latency and response time during classification. To evaluate performance under various activity conditions, we measured the time required for sensor data processing, classification, and output display, focusing on both dynamic and static activities.

The average **latency** of the system, defined as the time from data acquisition to classification output, was found to be approximately **120 ms**. This latency encompasses sensor data collection, preprocessing, and classification through the neural network model deployed on the STM32 microcontroller. This low latency is essential for applications requiring real-time feedback, such as health monitoring or activity tracking in elderly care, where timely response to activity changes is crucial.

The system's **response time**, defined as the interval between activity onset and recognition output, was measured across multiple trials for both dynamic and static activities. Results indicate an average response time of **150 ms** for dynamic activities, like walking, and a slightly higher response time of **200 ms** for static activities, such as sitting or standing. The increase in response time for static activities is attributed to the subtler sensor data variations, which necessitate more detailed analysis for accurate classification.

While the system demonstrates effective real-time responsiveness, the increased response time for static activities highlights a potential limitation in detecting subtle postural changes. Future optimizations in feature extraction and model tuning will aim to reduce response times further, especially for static activity recognition, enhancing the system's utility in applications where rapid detection of posture changes is critical.

5.4. Power Consumption Analysis

Power consumption is a critical consideration in embedded systems, particularly for wearable devices intended to operate continuously without frequent recharging. To evaluate the energy efficiency of our system, we conducted detailed measurements of the **average power consumption** of the STM32 microcontroller across various operational stages, including data acquisition, processing, and classification.

The system's average power consumption during **idle** periods (i.e., while awaiting sensor interrupts) was recorded at approximately **35 mW**. During **active data acquisition** and **real-time classification**, power usage rose to an average of **110 mW**. This efficiency is achieved through power management strategies, such as utilizing low-power modes during idle times and efficient scheduling of tasks, which minimize the active processing window.

These power metrics underscore the system's viability for deployment in battery-powered wearable devices, where achieving a balance between real-time responsiveness and energy efficiency is essential. By optimizing resource allocation and implementing quantization to reduce processing demands, our approach ensures the model remains effective within STM32's resource constraints while supporting continuous machine-learning inference with manageable power consumption.

Real-Time Inference Evaluation and Measurement

One of the critical aspects of deploying a Human Activity Recognition (HAR) system on embedded platforms is the system's ability to perform real-time inference. This section evaluates the real-time performance of the proposed HAR system by measuring latency, response time, and resource utilization during the inference phase.

1. Latency Measurement:

Latency is defined as the time taken between acquiring sensor data and producing a classification result. For real-time activity recognition, low latency is crucial to provide timely feedback in applications such as health monitoring or fitness tracking.

In our system, the average **inference latency** was measured at **120 ms**, encompassing sensor data acquisition, preprocessing, and neural network inference, optimized for the STM32 microcontroller using TensorFlow Lite Micro. Quantization techniques were employed to reduce the model size and computation requirements, enabling efficient inference within the microcontroller's limited resources. This low latency allows the system to respond almost instantaneously to user movements, which is essential for real-world applications where real-time feedback is critical.

2. Response Time:

The **response time** represents the duration between the occurrence of an activity and its recognition by the system, covering both processing latency and the time required for the sensors to capture the activity.

For dynamic activities (e.g., walking, running), the system achieved an average response time of **150 ms**, while more static activities (e.g., sitting, standing) showed a response time of **200 ms** due to subtler sensor data variations. This performance highlights the system's capability to provide rapid responses, essential for applications like fall detection in elderly care, where immediate feedback can be critical.

3. Resource Utilization and Power Consumption:

Real-time inference on embedded systems often faces challenges related to resource constraints and power consumption. Our system efficiently utilizes the available resources on the STM32 microcontroller, with the neural network occupying **45%** of the total memory available. The average power consumption during active data acquisition and classification was **110 mW**, ensuring long battery life in wearable applications.

This power-efficient design highlights the potential for deploying the HAR system in real-world, battery-powered devices where prolonged usage is critical.

6. Discussion

6.1. Efficacy and Challenges

Our model exhibits competitive performance in distinguishing dynamic activities, including walking, walking upstairs, and walking downstairs, achieving high precision and recall. This effectiveness is largely due to the model's architecture, which utilizes multimodal sensor fusion and refined feature extraction methods to capture distinctive motion patterns. When compared to similar HAR systems, such as those by David et al. [18] and Kopparapu et al. [19], our model demonstrates a latency advantage (120 ms compared to over 150 ms) and enhanced accuracy, particularly in dynamic activities, where it achieves 90% accuracy versus their 85–87%.

However, the model encounters challenges in static activity recognition, especially in differentiating sitting from standing. The subtle postural differences between these states are challenging for both feature extraction and sensor-based classification, as reflected in a relatively low recall for sitting (42%). This indicates a common misclassification with standing, suggesting the need for more sophisticated feature engineering or potentially additional sensors to better capture nuanced postures.

Addressing this limitation is crucial, especially for applications in healthcare and elderly monitoring, where accurate detection of both dynamic and static activities is essential for user safety and intervention. Future work may explore incorporating additional context-aware features or advanced sensor fusion techniques to improve static activity recognition, enhancing the system's overall reliability and efficacy for comprehensive activity monitoring.

6.2. Real-Time Classification Insights

Deploying our model on Tera Term for real-time classification provided valuable insights into its practical application, especially in scenarios where prompt and accurate activity recognition is essential, such as in elderly care for fall detection and in fitness monitoring for exercise quality. To better evaluate our model's performance, we conducted a comparative analysis with similar HAR systems, as shown in Table 8. This comparison demonstrates the model's effectiveness in achieving lower latency and higher classification accuracy, particularly for dynamic activities.

However, as with many HAR systems, recognizing static postures presents a challenge. Further refinement in sensor data processing and classification techniques will improve the model's utility in healthcare contexts that require precise differentiation between both dynamic and static activities. Table 8 provides a summary of key performance metrics compared with benchmarks from David et al. [18], Kopparapu et al. [19], and other recent studies.

Table 8. Comparative Performance Metrics for Human Activity Recognition Systems.

Metric	Our Model (STM32)	David et al. [18]	Kopparapu et al. [19]	Benchmark Avg.
Average Latency	120 ms	150 ms	170 ms	145 ms
Accuracy (Dynamic Activities)	90%	85%	87%	86%
Accuracy (Static Activities)	78.4%	72.6%	75%	74%
Power Consumption (Active)	110 mW	140 mW	130 mW	130 mW

6.3. Visual and Quantitative Analysis

The confusion matrix and various data plots—showcasing sensor readings and the model's learning progression over epochs—provide valuable quantitative and visual insights. These tools validate the model's adaptability and classification accuracy while also pinpointing specific areas needing enhancement, particularly in activities with closely related postures. The lower performance in static activity classification suggests that further development in feature extraction techniques or sensor fusion methods is necessary to improve model robustness in these scenarios, particularly for healthcare and eldercare applications.

6.4. Future Directions and Limitations

While the proposed HAR system demonstrates substantial potential, several limitations highlight areas for future improvement. Addressing these limitations, particularly regarding user variability and sensor noise, will enhance the model's robustness in diverse real-world applications such as healthcare, home monitoring, and outdoor activity tracking. Planned improvements and their expected benefits are summarized in Table 9, with a focus on individual variability in Table 10.

- **Impact of Sensor Noise:** The current model relies on low-pass filtering to manage sensor noise, particularly in MEMS data. While effective for high-frequency noise reduction, this method may inadvertently filter out subtle signals essential for static posture recognition (e.g., "Sitting" and "Standing"). This limitation affects the model's accuracy in low-motion scenarios. Future work will explore adaptive filtering techniques that dynamically adjust based on activity type, thus balancing noise reduction and data preservation.
- **User Variability and Physical Characteristics**: The model does not currently account for variations in user height, weight, or mobility, which influence acceleration and sensor readings. Such variability can reduce accuracy, especially across diverse user populations. Future work will expand the dataset to include participants with different physical attributes and implement synthetic data generation to simulate various body types. Transfer learning techniques will also be explored to enable model adaptation to individual differences, as shown in Table 10.
- **Low F1-Scores for Static Activities:** The model exhibits lower F1-scores for static activities, particularly "Sitting" and "Standing". This limitation is critical for applications

in elder monitoring and health tracking. Future efforts will focus on advanced feature extraction, such as spectral and time-frequency analysis, to better differentiate static postures. Additionally, comparisons with complex deep-learning architectures (e.g., CNN-LSTM hybrids) will be conducted to evaluate potential model improvements for static activity recognition.

- **Adaptive Filtering and Noise Management:** To address limitations posed by fixed filtering, future iterations will incorporate adaptive filtering techniques that adjust dynamically based on detected activity. This approach is intended to balance noise reduction while preserving essential data for static activities, improving accuracy in low-motion scenarios, and enhancing robustness across varied environments.
- **Synthetic Data Augmentation for Robustness**: Employing synthetic data generation will increase training sample diversity, especially for underrepresented static activities. This method will support generalization across controlled and uncontrolled environments, allowing the model to adapt to variations in sensor alignment, individual movement patterns, and environmental noise, thus enhancing robustness in real-world applications.
- **Algorithm and Model Optimization:** Exploring advanced neural architectures, such as transformer-based models or hybrid CNN-RNN structures, may improve classification accuracy across both static and dynamic activities. Future work will prioritize resilience to noise and sensor misalignments to ensure reliable performance in resource-constrained environments.
- **Enhanced Sensor Fusion:** Integrating additional sensors, such as gyroscopes, magnetometers, or environmental sensors, can expand the feature set and aid in differentiating between similar static activities. Multimodal data fusion will facilitate activity recognition across diverse contexts, improving model adaptability under uncontrolled conditions.

The current model's success in dynamic activity recognition underscores its practical potential. However, addressing limitations related to sensor noise, user variability, and static posture recognition is crucial for healthcare and personalized fitness applications. By improving adaptability, noise resilience, and user-specific responsiveness, future advancements will contribute to a reliable and accessible HAR system, supporting quality-of-life improvements through advanced, personalized monitoring.

Table 9. Planned Future Improvements and Expected Benefits.

Improvement Area	Goal	Expected Benefit
Incorporating Individual Variability	Increase robustness across diverse user groups	Improved generalization and reduced bias across populations
Advanced Feature Engineering	Improve static activity recognition	Higher accuracy in distinguishing subtle postures
Adaptive Filtering and Noise Management	Dynamic noise management based on activity type	Enhanced accuracy in low-motion and noisy environments
Synthetic Data Augmentation	Enhance generalization across environments	Better performance in uncontrolled settings
Algorithm Optimization	Increase classification resilience to noise	Robustness in noisy and varied environments
Enhanced Sensor Fusion	Integrate additional sensor modalities	Improved activity differentiation across contexts

Table 10. Expected Impact of Incorporating Individual Variability on Model Performance.

Aspect of Variability	Potential Model Adjustment	Expected Benefit
Age Differences	Age-specific tuning or synthetic data	Improved recognition for older adults
Body Type (Height/Weight)	Synthetic data reflecting varied body types	Enhanced accuracy across diverse body types
Physical Mobility Levels	Transfer learning for limited mobility data	Higher accuracy in healthcare applications

6.5. Impact of Low-Pass Filtering on Classification Accuracy

The use of a low-pass filter plays a crucial role in reducing sensor noise; however, it also poses the risk of removing minute details in sensor data that are essential for classifying static or subtle activities. We observed that while the filter enhanced performance in dynamic activity classification, it introduced some limitations in recognizing static activities like sitting and standing, which often involve minimal movement.

To address this, future iterations of our model will explore adaptive filtering or a hybrid approach, combining low-pass filtering for dynamic activities with alternative methods, such as threshold-based filtering, for static activities. This approach aims to retain high-frequency details that are potentially useful for distinguishing subtle postures and, therefore, improving classification accuracy across both activity types.

6.6. Limitations of Uniform Movement Assumption

The current model operates under the assumption that all individuals exhibit uniform movement patterns. This simplification overlooks inter-individual variability that can arise due to differences in age, body weight, height, and physical condition. Such factors can influence movement dynamics, potentially affecting classification accuracy across different user groups. For instance, older adults or individuals with limited mobility may display distinct movement characteristics, which may lead to misclassification in real-world applications.

To address this limitation, future studies will focus on incorporating user-specific factors and variability-aware training methods. This could involve generating synthetic data that reflects diverse body types and ages or employing transfer learning techniques to adapt the model to different physical characteristics. Integrating such approaches could enhance the model's robustness across diverse populations, making it more applicable to real-world settings.

7. Conclusions

This research has advanced the field of human activity recognition by developing a robust model with high precision and recall in identifying dynamic activities. The model's strong performance in recognizing activities like walking, walking upstairs, and walking downstairs highlights the effectiveness of our sensor integration and feature extraction approaches. Additionally, the model's ability to perform real-time classification via Tera Term reinforces its practicality for deployment in everyday technology, including wearable and mobile applications.

However, challenges remain in accurately distinguishing between static activities such as sitting and standing, where subtle postural nuances lead to lower recall rates. The model's lower performance for sitting reflects a need for further refinement in feature extraction methods and potential benefits from incorporating additional sensor data. This enhancement is crucial for applications in healthcare, particularly in elderly monitoring and fall detection, where precise differentiation between static and dynamic activities is critical.

Looking ahead, future work will focus on key improvements such as advanced feature engineering, data augmentation, optimized machine-learning algorithms, and sensor fusion. These enhancements are expected to address current limitations by increasing the diversity of training data and enriching the feature set, therefore improving the model's ability to discern closely related activities, including static postures.

Ongoing development in this area has significant potential for a wide range of applications, from health monitoring and elderly care to fitness tracking and sports science. Such improvements will enhance the accuracy of activity recognition and expand device capabilities to provide more personalized and context-aware health insights. Ultimately, by refining our model and broadening its real-world applicability, we aim to contribute to the development of innovative solutions that support quality of life and promote better health outcomes through precise, real-time understanding of human activities.

Author Contributions: P.T., B.A., and F.M. contributed equally to the conceptualization, methodology, implementation, formal analysis, and manuscript preparation. All authors participated in the design and development of the human activity recognition system, as well as the drafting and critical revision of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the FCT – Fundação para a Ciência e a Tecnologia, I.P. [Project UIDB/05105/2020].

Data Availability Statement: Data Used for this work is available at https://www.kaggle.com/datasets/pardhut/pardhu-har-data, (accessed on 29 September 2024).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bulling, A.; Blanke, U.; Schiele, B. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Comput. Surv.* **2014**, *46*, 1–33. [CrossRef]
2. Vaizman, Y.; Ellis, K.; Lanckriet, G. Context recognition in-the-wild: Unified model for multi-modal sensors and multi-label classification. Proc. ACM Interactive Mobile Wearable Ubiquitous Technol. **2018**, *1*, 1–22. [CrossRef]
3. Lara, O.D.; Labrador, M.A. A survey on human activity recognition using wearable sensors. *IEEE Commun. Surv. Tutorials* **2012**, *15*, 1192–1209. [CrossRef]
4. Kwapisz, J.R.; Weiss, G.M.; Moore, S.A. Activity recognition using cell phone accelerometers. In Proceedings of the 4th International Workshop on Knowledge Discovery from Sensor Data, Washington, DC, USA, 25–28 July 2010; pp. 10–18.
5. Ignatov, A. Real-time human activity recognition from accelerometer data using convolutional neural networks. *Appl. Soft Comput.* **2018**, *62*, 915–922. [CrossRef]
6. Wang, J.; Chen, Y.; Hao, S.; Peng, X.; Hu, L. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognit. Lett.* **2019**, *119*, 3–11. [CrossRef]
7. Ordóñez, F.J.; Roggen, D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors* **2016**, *16*, 115. [CrossRef]
8. Ravi, N.; Dandekar, N.; Mysore, P.; Littman, M. Activity recognition from accelerometer data. In *Proceedings of the National Conference on Artificial Intelligence*; AAAI: Washington, DC, USA, 2005.
9. Yurtman, A.; Barshan, B. Human activity recognition from wearable IMU using dominant feature sets. *J. Biomed. Inform.* **2014**, *47*, 42–53.
10. Ha, S.; Choi, S.; Lee, S.; Lee, H.J. Multi-modal convolutional neural networks for activity recognition. In Proceedings of the 2015 IEEE International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland, 12–17 July 2015; pp. 907–910.
11. TinyML. Available online: https://www.tinyml.org (accessed on 10 March 2023).
12. Guan, J.; Plötz, T. Ensembles of deep LSTM learners for activity recognition using wearables. In Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, Maui, HI, USA, 11–12 September 2017; pp. 157–160.
13. Shotton, E.; Fitzgibbon, A.; Cook, M.; Sharp, T.; Finocchio, M.; Moore, R.; Kipman, A.; Blake, A. Real-time human pose recognition in parts from a single depth image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 1297–1304.
14. Huang, K.; Hu, W.; Feng, S.R. Human skeletal tracking using Kinect. *J. Robot.* **2015**, *2015*, 782523.
15. Jamil, H.; Qayyum, F.; Iqbal, N.; Jamil, F. Optimal ensemble scheme for human activity recognition and floor detection based on AutoML and weighted soft voting using smartphone sensors. *IEEE Sens. J.* **2022**, *22*, 13213–13224. [CrossRef]
16. Diraco, G.; Rescio, G.; Siciliano, P.; Leone, A. Review on human action recognition in smart living: Sensing technology, multimodality, real-time processing, interoperability, and resource-constrained processing. *Sensors* **2023**, *23*, 5281. [CrossRef]
17. Artificial Intelligence (AI) Software Expansion for STM32Cube. STMicroelectronics. 2023. Available online: https://www.st.com/en/embedded-software/x-cube-ai.html (accessed on 3 October 2024).
18. David, R.; Duke, J.; Jain, A.; Reddi, V.J.; Jeffries, N.; Li, J.; Kreeger, N.; Nappier, I.; Natraj, M. TensorFlow Lite Micro: Embedded machine learning on TinyML systems. *arXiv* **2020**, arXiv:2010.08678.

19. Kopparapu, K.; Lin, E.; Breslin, J.G.; Sudharsan, B. TinyFedTL: Federated transfer learning on ubiquitous Tiny IoT devices. In Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events, Pisa, Italy, 20–24 March 2022; pp. 79–81.
20. McMahan, H.B.; Moore, E.; Ramage, D.; Hampson, S.; Arcas, B.A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, Ft. Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.
21. Barr, M.; Massa, A. *Programming Embedded Systems in C and C++*; O'Reilly Media: Newton, MA, USA, 2006.
22. Sena, J.; Barreto, J.; Caetano, C.; Schwartz, W.R. Human activity recognition based on smartphone and wearable sensors using multiscale DCNN ensemble. *Neurocomputing* **2021**, *421*, 114–126. [CrossRef]
23. Muaaz, M.; Chelli, A.; Pätzold, M. Wi-Sense: A passive human activity recognition system using Wi-Fi and convolutional neural network and its integration in health information systems. *Ann. Telecommun.* **2022**, *77*, 335–353. [CrossRef]
24. Cai, H.; Gan, C.; Zhu, L.; Han, S. TinyTL: Reduce memory, not parameters for efficient on-device learning. In Proceedings of the International Conference on Advances in Neural Information Processing Systems, Virtual, 6–12 December 2020; Volume 33, pp. 11285–11297.
25. Rodrigues, V.F.; Bazo, R.; Reis, E.; Seewald, L.A. Baptizo: A sensor fusion based model for tracking the identity of human poses. *Inf. Fusion* **2020**, *62*, 165–178.
26. Negi, A.; Kumar, K. End-to-end residual learning-based deep neural network model deployment for human activity recognition. *Appl. Intell.* **2023**, *53*, 2753–2767. [CrossRef]
27. Lin, J.; Zhu, L.; Chen, W.-M.; Wang, W.-C.; Gan, C.; Han, S. On-device training under 256 KB memory. *arXiv* **2022**, arXiv:2206.15472.
28. Sudharsan, B.; Yadav, P.; Breslin, J.G.; Ali, M.I. Train++: An incremental ML model training algorithm to create self-learning IoT devices. In Proceedings of the 18th IEEE International Conference on Ubiquitous Intelligence and Computing, Guangzhou, China, 29 December–1 January 2021; pp. 97–106.
29. Angerbauer, S.; Palmanshofer, A.; Selinger, S.; Kurz, M. Comparing human activity recognition models based on complexity and resource usage. *Appl. Sci.* **2021**, *11*, 8473. [CrossRef]
30. Liu, Y.; Yang, Z.; Chen, W. Human activity recognition with LSTM model on accelerometer data. *J. Sens. Technol.* **2020**, *9*, 23–32.
31. Yi, M.-K.; Hwang, S.O. Smartphone based human activity recognition using 1D lightweight convolutional neural network. In Proceedings of the International Conference on Electronics, Information, and Communication, Jeju, Republic of Korea, 6–9 February 2022; pp. 1–3.
32. Human Activity Recognition Using CNN in Keras for Sensortile. Available online: https://github.com/ausilianapoli/HAR-CNN-Keras-STM32/blob/master/Dataset.csv (accessed on 1 February 2023).
33. Zhu, C.; Sheng, W.; Liu, M. Wearable sensor-based human activity recognition with hybrid classifier models. *IEEE Trans. -Hum.-Mach. Syst.* **2016**, *46*, 769–779
34. Zeng, M.; Nguyen, L.T.; Yu, B.; Mengshoel, O.J.; Zhu, J.; Wu, P.; Zhang, J. Convolutional neural networks for human activity recognition using mobile sensors. In Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, WA, USA, 13–17 September 2014; pp. 197–200.
35. Warden, P.; Situnayake, D. *TinyML: Machine Learning with TensorFlow on Arduino and Ultra-Low-Power Microcontrollers*; O'Reilly Media: Newton, MA, USA, 2019.
36. Mekruksavanich, S.; Jantawong, P. A lightweight deep convolutional neural network with squeeze-and-excitation modules for efficient human activity recognition using smartphone sensors. *Proc. IEEE Big Data Anal.* **2021**, *18*, 1007–1015.
37. Vakili, M.; Rezaei, M. Incremental learning techniques for online human activity recognition. *arXiv* **2021**, arXiv:2109.09435.
38. Gaya-Morey, F.X.; Buades-Rubio, J.M.; Manresa-Yee, C. Deep learning for computer vision-based activity recognition and fall detection of the elderly: A systematic review. *Appl. Intell.* **2024**, *54*, 2145–2162. [CrossRef]
39. Riahi, M.; Eslami, M.; Safavi, S.H. Human activity recognition using improved dynamic image. *IET Image Process.* **2020**, *14*, 735–746. [CrossRef]
40. Malekmohamadi, H.; Pattanjak, N.; Bom, R. Human activity identification in smart daily environments. In *Smart Assisted Living*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 49–65.
41. Ramos, R.G.; Domingo, J.D.; Zalama, E. Daily human activity recognition using non-intrusive sensors. *Sensors* **2021**, *21*, 5270. [CrossRef] [PubMed]
42. Kalabakov, S.; Stankoski, S.; Kiprijanovska, I.; Andova, A. What actually works for activity recognition in scenarios with significant domain shift: Lessons learned from the 2019 and 2020 Sussex-Huawei challenges. *Sensors* **2022**, *22*, 3613. [CrossRef]
43. Wang, L.; Huynh, D.Q.; Koniusz, P. A comparative review of recent kinect-based action recognition algorithms. *IEEE Trans. Image Process.* **2019**, *28*, 4367–4383. [CrossRef]