CHAPTER-1

INTRODUCTION

1.1 Thesis Abstract

Human activity recognition (HAR) is a critical component of various applications, including healthcare monitoring, sports analytics, and human-computer interaction. In recent years, inertial sensors embedded in wearable devices have gained popularity for their ability to capture rich spatiotemporal information related to human movements. This thesis presents an in-depth exploration of spatial-temporal analysis techniques applied to HAR using inertial sensors. The dissertation begins with a comprehensive overview of the HAR domain, highlighting the importance of spatial-temporal features in achieving accurate recognition. It also discusses the various challenges and limitations associated with traditional methods and emphasizes the need for advanced techniques to enhance HAR performance.

This thesis explores the design and development of novel frameworks for HAR, focusing on the utilization of inertial sensors, such as accelerometers and gyroscopes. The framework integrates preprocessing steps, and machine learning algorithms to capture the intricate spatial-temporal patterns embedded within the sensor data.

This thesis contains five chapters, which have been organized briefly to explore the advancement in classifying different activities using the smartphone sensor data. Each chapter begins with a brief description of its contents followed by motivation for the research, and further are supported by graphical illustrations.

The thesis briefly discusses the evolution of HAR, the basic working mechanism for HAR, various recognition techniques, objective and challenges in the HAR in ***Chapter 1*****.**

Chapter 2 outlines various architectures, evolution of HAR, literature survey/motivation behind HAR, sensor modality, concept of eXplainability and attention mechanism followed by the dataset description, framework and evaluation metrics.

Chapter 3**,** deals with use of a segmentation model for the task of HAR. This chapter explores the introduction of a classifying block which is used to identify different actions performed by humans using the inbuilt smartphone sensors. It has been found that the introduction of the classifying block in the segmentation model can effectively classify normal as well as transition activities.

In ***Chapter 4*** an explainable and cost-efficient self-attention-based CNN-LSTM model has been presented which is capable of classifying different activities. This chapter envisages a use of self-attention in the CNN-LSTM based architecture which helps in recognizing different activities. The self-attention layer in the CNN-LSTM architecture focuses on the relevant features responsible for recognizing different activities. Furthermore, the 1-D Grad CAM was implemented to study the black box nature of DL models. In this chapter we also performed a network study which justifies the used of each layer in the network.

Chapter 5 presents a summary of the outcomes of the study undertaken during the work. The overall conclusion drawn from the study has been enunciated. The study revealed that it is possible to recognize different activities using smartphone data. It can also be concluded that with the suitable selection of different layers in the DL models can help in building a cost-efficient model. Further, the use of 1-D-GradCAM can unbox the black box nature of these DL models. This chapter further outlines the future scope of this extensive study.

1.2 Introduction on HAR

Human activity recognition, also known as human behavior analysis, is the process of using technology to identify and understand human behavior. This can be done through the use of various sensors and data-gathering devices, such as cameras and wearable devices, that collect information on a person's movements, interactions, and environment.

One of the main applications of human activity recognition is in the field of surveillance, where it can be used to monitor public spaces and identify potential security threats. It can also be used in the design and optimization of buildings and public spaces, as well as in the field of healthcare, where it can be used to monitor and support the care of patients with mobility or cognitive impairments.

Recently, with the increasing popularity of IoT and other smart devices, human activity recognition has also been used in smart homes to automate various actions and tasks based on a person's habits and routines. This can include controlling lighting, temperature, and appliances, as well as providing assistance with tasks such as cooking and cleaning.

Despite its many potential uses, human activity recognition also raises concerns about privacy and security. The use of data-gathering devices and sensors can collect sensitive information about a person's movements and behavior, and there are concerns about how this information is used and stored. Additionally, the use of human activity recognition in public spaces can raise concerns about government surveillance and the potential for abuse of the technology.

To address these concerns, it is important to have clear guidelines and regulations in place to govern the use of human activity recognition. This can include measures to ensure that data is collected and used in a transparent and secure manner, as well as to protect individuals' privacy.

In conclusion, human activity recognition is a powerful technology that has many potential uses in various fields such as surveillance, healthcare, and smart homes. However, it also raises important concerns about privacy and security, which need to be addressed through clear guidelines and regulations.

One of the key techniques used in human activity recognition is computer vision, which involves using cameras and other imaging devices to capture and analyze visual data. This can include tracking a person's movements and gestures, as well as recognizing objects and other features in the environment. Computer vision algorithms can also be used to extract features such as color, texture, and shape, which can be used to identify and classify different activities.

Another important technique used in human activity recognition is machine learning. Machine learning algorithms can be trained on large datasets of labeled activity data to learn patterns and features that are associated with different activities. This allows the system to recognize and classify new activities based on the patterns and features it has learned.

In addition to computer vision and machine learning, human activity recognition also makes use of other technologies such as sensors and wireless communication. Wearable devices, for example, can be used to gather data on a person's movements, heart rate, and other physiological markers, which can provide additional information about their activity and behavior.

As for the application in healthcare, human activity recognition can be used to monitor patients with mobility or cognitive impairments, for example, to detect falls, monitor sleep patterns, or track medication adherence. It can also be used to monitor patients with chronic conditions such as diabetes, to track their physical activity and help manage their condition.

Another important application of human activity recognition is in the field of sports and fitness, where it can be used to track and analyze an athlete's performance, identify areas for improvement, and optimize training programs. Additionally, it is also being used in the field of gaming and entertainment, to create more immersive and interactive experiences.

In terms of privacy and security, there are a number of issues that need to be considered when using human activity recognition. These include data protection and privacy, which involves ensuring that the data collected is used in a secure and responsible manner, and that individuals' personal information is protected. Additionally, there are ethical concerns about the use of human activity recognition in public spaces, including issues related to surveillance, profiling, and discrimination.

To address these issues, it is important to have clear guidelines and regulations in place to govern the use of human activity recognition. This can include measures to ensure that data is collected and used in a transparent and secure manner, as well as to protect individuals' privacy.

1.3 Evolution of HAR

The evolution of Human Activity Recognition (HAR) has undergone significant advancements over the years, driven by technological innovation and the growing demand for applications that understand and respond to human behavior. Here's a chronological overview of the key stages in the evolution of HAR:

- **Early Approaches (Pre-2000s):** In the pre-2000s, the field of HAR was in its initial stages, marked by early approaches that laid the foundation for future developments. Basic sensor technology, primarily accelerometers and gyroscopes, formed the core of HAR systems, capturing motion data from the body or objects. Researchers relied on handcrafted feature engineering to represent different aspects of human activities. These early systems often used rule-based methods and focused on recognizing simpler activities like walking and running, with applications primarily in healthcare and biomechanics. However, the early HAR systems had limitations in terms of accuracy and adaptability to complex, real-world activities. The concept was first introduced and explored by the Neural Network house [1] in the field of home automation, as well as in various location-based apps that seek to customize systems based on users' whereabouts [2] [3]. The approach was quickly recognized as being highly valuable and appropriate within the domain of ubiquitous and mobile computing, which was an emerging field in the late 1990s. This recognition was primarily driven by the approach's ease of implementation. Therefore, a significant amount of research has been conducted to explore the utilization of sensors in different contexts of ubiquitous and mobile computing. This has resulted in substantial efforts in the areas of context awareness [4]–[6], smart appliances [7], [8], and activity detection [9]–[12]. During that period, a majority of studies involved wearable sensors, which included either specialized sensors attached to individuals or portable devices such mobile devices. These technologies were mostly utilized in ubiquitous computing contexts, particularly in facilitating the provision of context-awareness for mobile devices.

- **Feature Engineering and Machine Learning (2000s):** In the early 2000s, there was indeed a growing interest in sensor-based approaches for monitoring human activities, and this approach complemented the application of machine learning techniques to Human Activity Recognition (HAR) such as decision trees, support vector machines, and Hidden Markov Models (HMMs), started to be applied to HAR. The sensor-based approach involved attaching sensors to objects or wearable devices that individuals interacted with during their activities. These sensors could include accelerometers, gyroscopes, magnetometers, and other types of sensors capable of capturing motion and environmental data. Feature engineering played a crucial role in this context. It involved the extraction of meaningful features from the sensor data that could be used as input for machine learning models. These features were designed to capture relevant characteristics of human activities and were essential for accurate classification.

- **Introduction of Smartphones and Wearables (2010s):** The proliferation of smartphones and wearable devices with built-in sensors revolutionized HAR. The use of accelerometers, gyroscopes, and GPS in these devices allowed for more accurate and fine-grained activity recognition. Machine learning algorithms like Random Forests, k-Nearest Neighbors, and Neural Networks gained traction.
- **Deep Learning and Sensor Fusion (Mid-2010s):** Deep learning techniques, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), began to dominate HAR research. These models could automatically learn features from raw sensor data, reducing the need for manual feature engineering. Sensor fusion, combining data from multiple sensors, led to improved accuracy.
- **Temporal Modeling and Recurrent Networks (Late 2010s):** HAR evolved to consider temporal relationships between sensor readings. Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) gained popularity for sequence modeling. This enabled recognition of activities with intricate temporal dependencies.
- **Attention Mechanisms and Transformers (Late 2010s - Early 2020s):** Attention mechanisms and Transformers, initially popular in natural language processing,

found applications in HAR. These mechanisms enabled the models to focus on relevant parts of the data while considering contextual information. Self-attention and multi-head attention mechanisms were integrated into HAR architectures.

- **Real-World Applications and Commercialization (2020s):** HAR moved from research laboratories to real-world applications. Smartwatches, fitness trackers, and health monitoring devices integrated HAR for tracking physical activities and providing personalized recommendations. Industrial settings adopted HAR for workplace safety, and security systems integrated activity recognition for surveillance.
- **Privacy and Ethics (2020s):** With the increasing use of sensor data and HAR in various applications, concerns about privacy, data security, and ethics emerged. Researchers and practitioners began addressing these issues to ensure responsible development and deployment of HAR systems.
- **Multi-Modal and Few-Shot Learning (Present and Beyond):** Current research is exploring multi-modal approaches that fuse data from various sensors and modalities for more comprehensive activity recognition. Few-shot learning techniques are being developed to train models with minimal labeled data, expanding the scope of HAR to new activities.
- **Explainable AI and Cross-Domain Applications (Future):** The future of HAR involves making AI models more interpretable and explainable. HAR techniques developed in one domain, such as healthcare, might find applications in other domains, like robotics or sports analysis.

The evolution of HAR reflects the interplay between advances in sensor technology, machine learning algorithms, and the increasing integration of AI into our daily lives. As technology continues to advance, HAR will likely become even more accurate, adaptable, and ubiquitous across a wide range of applications.

1.4 Motivation

Human activity recognition (HAR) using wearable sensors is an active research area because it has the potential to impact many areas, including healthcare, sports, and entertainment. Some of the motivations for human activity recognition using wearable sensors are:

- Healthcare: Wearable sensors can be used to monitor patients with chronic conditions such as diabetes, heart disease, and obesity. They can provide real-time data on physical activity, heart rate, and other vital signs, which can be used to manage the condition and prevent complications.
- Sports: Wearable sensors can be used to track the performance of athletes and help coaches to optimize training programs. They can also be used to monitor fatigue, injury risk, and recovery.
- Entertainment: Wearable sensors can be used to create new interactive experiences, such as gaming and virtual reality.
- Quality of life: Wearable sensors can be used to improve the quality of life for older adults, people with disabilities, and those who live alone by monitoring their physical activity, fall detection, and providing early warning of potential health issues.
- Safety and security: Wearable sensors can be used to monitor the safety of workers, such as construction workers, miners, and firefighters, and provide early warning of dangerous conditions. They can also be used for security purposes, such as tracking people in crowded areas or monitoring the activity of people in high-risk situations.

Thus, we can say that human activity recognition using wearable sensors has the potential to improve healthcare, sports, entertainment, quality of life, and safety and security.

1.5 Objective of HAR

The main objective of HAR is to automatically recognize and classify the activities performed by a person based on sensor data or visual data. HAR is a technique that involves using machine learning algorithms to analyze the data collected from sensors or cameras, and to identify patterns that correspond to specific activities.

The objective of HAR can vary depending on the specific application. Some of the common objectives of HAR include:

- **Health monitoring:** In healthcare, HAR can be used to monitor the physical activity of patients, track their progress, and provide feedback to healthcare professionals. The objective is to improve patient outcomes and prevent complications.
- **Sports performance monitoring:** In sports, HAR can be used to measure the physical activity of athletes during training and competition, track their progress, and provide feedback to coaches. The objective is to optimize athlete performance and prevent injuries.
- **Robotics:** In robotics, HAR can be used to enable more sophisticated human-robot interaction. The objective is to improve the accuracy of robot actions and make them more responsive to human behavior.
- **Surveillance:** In surveillance, HAR can be used to detect and prevent criminal activity, detect potential safety hazards in industrial environments, and more. The objective is to improve public safety and prevent criminal activity.

1.6 Problem Statement

Human Activity Recognition (HAR) is a multidisciplinary field that intersects sensor technology, machine learning, and real-world applications. It plays a crucial role in various fields such as healthcare, sports analytics, and human-computer interaction. The accurate classification of human activities from sensor data remains a challenging task due to the complexity and variability of human movements. Recent advancements in deep learning have shown promising results in addressing this challenge. However, there are still limitations in existing approaches, including computational complexity, interpretability of models, and optimization of network architecture.

1.6.1. The Need for Accurate HAR:

Human activity recognition plays a pivotal role in diverse fields, such as healthcare, sports, security, and smart environments, where understanding and predicting human behavior are of paramount importance. In each of these domains, the accuracy of HAR directly impacts the effectiveness and reliability of systems and applications.

1.6.2. Limitations in Current Approaches:

Understanding human activities from sensor data is essential for numerous applications across domains such as healthcare and sports analytics. However, this task poses significant challenges due to the inherent complexity and variability of human movements. Despite advancements in deep learning techniques, several persistent issues hinder the effectiveness of current approaches. One major challenge is the complexity of the models themselves, which can be overly intricate and difficult to interpret. Additionally, these models often struggle to efficiently process the vast amounts of data generated by sensor inputs, leading to scalability issues and computational inefficiencies. Moreover, the lack of interpretability in many deep learning models poses a substantial barrier to understanding how and why decisions are made, limiting their practical utility in real-world scenarios. Consequently, it is imporatnt to address these limitations and develop more effective and interpretable deep learning frameworks for HAR.

1.6.3. Objective of the thesis:

The primary objective of this thesis is to develop an effective and interpretable deep learning framework for HAR that excels in both performance and efficiency. This framework aims to utilize the strengths of novel architectural designs and methodological innovations to achieve state-of-the-art results in HAR while providing insights into the underlying features contributing to activity recognition. By integrating innovative architectural designs with comprehensive model analysis and interpretability techniques, this thesis aims to address the limitations present in existing HAR models. The goal is to enhance both the performance and interpretability of HAR models, thereby advancing the state-of-the-art in the field. Through a holistic approach that combines architectural innovations with methodological advancements, this research endeavors to develop a deep learning framework that not only achieves superior performance in HAR tasks but also provides valuable insights into the underlying mechanisms driving activity recognition.

1.6.4. Approach:

This research thesis proposes a synergistic approach that combines the contributions of two deep learning model (1) the novel CNN-LSTM self-attention architecture for HAR, and (2) the Seq2Dense U-Net architecture for HAR. The former introduces a deep learning architecture that leverages convolutional neural networks (CNNs) and long short-term memory (LSTM) networks with self-attention mechanisms for HAR. This architecture demonstrates improved performance in terms of classification accuracy while also addressing computational efficiency through FLOPS analysis and model depth optimization. The later focuses on enhancing the interpretability of HAR models by the incorporation of 1-D Gradient-weighted Class Activation Mapping (Grad-CAM) enhances the interpretability of the model by elucidating the salient features contributing to activity classification. Additionally, this thesis explores the depth analysis of the network to determine the optimal architecture for activity recognition.

1.6.5. Significance of the work:

By combining the strengths of both approaches, this research aims to develop a comprehensive framework that not only achieves state-of-the-art performance in HAR but also provides insights into the underlying features contributing to activity classification. The proposed framework has the potential to impact various applications, including healthcare monitoring, sports performance analysis, and assistive technologies for individuals with motor impairments.

1.7 Outlines of the Thesis

The thesis aims to extensive study of HAR using smartphone data. This study has attempted to resolve the complexity in recognizing the different activities using a simple, cost-effective model. With the advancement in the DL technologies and techniques, this thesis also includes the use of self-attention which only focuses on the relevant information thus making the model simple and cost effective. This thesis also includes the explainability of the model which demonstrates which features are responsible for the particular activity.

The thesis is divided into five chapters to cover the extensive investigation of the HAR during my Ph.D. program. The thesis covers the introduction section, the analysis of the sequential data for HAR, and the spatial temporal analysis of the inertial sensor data which also includes the explainability of the model. The thesis chapters are organized as follows-

Chapter 1 presents an in-depth investigation, spanning multiple key facets pertaining to HAR. The chapter begins with a concise and informative abstract that presents a brief overview of the main emphasis and contributions of the subsequent research. Following this, a comprehensive overview of HAR is provided, explaining the field's fundamental importance, exploring the evolution of HAR, following the historical development and advancements of this technical field. In addition, it illustrates the fundamental reasons behind the research, explores the specific goals to be accomplished, and formulates the problem statement that the study aims to resolve. Finally, it concludes by providing an outline of the thesis, providing readers with a roadmap for the subsequent chapters and the key areas of focus within the research.

Chapter 2 presents a comprehensive overview of the multifaceted field of HAR covering essential aspects such as types of HAR, sensor modality, emphasizing the pivotal role of different sensor types in capturing data for HAR systems. The process of HAR is outlined, a detailed explanation of the sequential stages involved in recognizing and categorizing human activities is also provided. The chapter also provides an examination of the state-of-the-art methods in HAR, discussing the current landscape of techniques and approaches. Furthermore, explainability concepts, attention mechanism, data, algorithms, frameworks, and evaluation metrics are also discussed.

Chapter 3 introduces the concept of employing a segmentation model for classification of various activities. The significance of sequential data in the field of HAR is detailed in this chapter. The key highlights of this chapter include an introductory section, where we set the stage for the exploration of segmentation based HAR and introduces the concept of using a segmentation model for classification of various activities. The importance of sequential data for the HAR is also discussed. Further, the novel Seq2Dense U-Net architecture, and its implementation is covered. Following that, results of experiments performed using the proposed model are presented and outcomes are detailed. Network depth analysis is a crucial section, where the complexities of the design are thoroughly examined, and its impact is evaluated. Finally, the chapter concludes, summarizing the findings and insights gained from the utilization of the segmentation model for HAR classification.

Chapter 4 covers the significance and practical applications of self-attention within a CNN-LSTM based architecture. The chapter focuses on the development of a cost-efficient model using self-attention, shedding light on the novel Self-attention-based CNN-LSTM architecture. The research findings are highlighted in this chapter, with a detailed explanation of the results obtained with the proposed model. Additionally, this chapter presents the realm of model explainability by implementing a 1-D GradCAM approach, offering insights into how our model arrives at its decisions. Network analysis is a key section where we dissect the architecture's performance and intricacies. The chapter concludes by summarizing the research outcomes and the insights gained through the application of self-attention in the CNN-LSTM architecture.

Chapter 5 concludes the summary and future scope of the research work presented in the previous chapters.