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