DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

APPLICATION OF DEEP LEARNING METHODS IN HUMAN ACTIVITY RECOGNITION

by
Kemal BAYSARI

October 2022
İZMİR

APPLICATION OF DEEP LEARNING METHODS IN HUMAN ACTIVITY RECOGNITION

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the MSc. degree in Computer Engineering

by
Kemal BAYSARI

October 2022
İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **“APPLICATION OF DEEP LEARNING METHODS IN HUMAN ACTIVITY RECOGNITION”** completed by **KEMAL BAYSARI** under the supervision of **ASSOC. PROF. DR. DERYA BİRANT** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Derya BİRANT

Supervisor

Dr. Öğr. Üyesi Erdogan CAMCIOĞLU

(Jury Member)

Prof. Dr. Alp KUT

(Jury Member)

Prof. Dr. Okan FISTIKOĞLU

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

I want to thank my supervisor, Assoc. Prof. Dr. Derya BİRANT, for her immense knowledge, great and continuous support, magnificent mentorship, encouragement, and useful suggestions for this thesis. She helped me all the time with the research and writing of this thesis. It has been a good opportunity to be her MSc. student.

Also, I want like to thank my father, mother, and my wife for their support.

Kemal BAYSARI

APPLICATION OF DEEP LEARNING METHODS IN HUMAN ACTIVITY RECOGNITION

ABSTRACT

Traditional sensor-based human activity recognition (HAR) has been defined as a time-series data classification problem and requires feature extraction. The current HAR systems still lack transparent, interpretable, and explainable approaches that can generate human-understandable information. This thesis proposes an approach, called Human Activity Recognition on Signal Images (HARSI), which defines the HAR problem as an image classification problem to improve both explainability and recognition accuracy. The proposed HARSI approach transforms the smart sensor data into some visual images to take advantage of the strengths of convolutional neural networks (CNN) in handling image data. The experimental results carried out on a real-world dataset showed that a significant improvement was achieved by the proposed HARSI model compared to the traditional machine learning models. The results also showed that our method outperformed the state-of-the-art methods in terms of classification accuracy.

Keywords: Human activity recognition, machine learning, deep learning, smart sensor systems, explainable artificial intelligence.

İNSAN AKTİVİTESİ TANIMADA DERİN ÖĞRENME YÖNTEMLERİNİN UYGULANMASI

ÖZ

Geleneksel sensör tabanlı insan aktivitesinin tanınması (HAR), bir zaman serisi verisi sınıflandırma problemi olarak tanımlanmıştır ve özellik çıkarımı gerektirir. Mevcut HAR sistemleri, insan tarafından anlaşılabilir bilgiler üretebilen şeffaf, yorumlanabilir ve açıklanabilir yaklaşımlardan hala yoksundur. Bu tez, hem açıklanabilirliği hem de tanıma doğruluğunu geliştirmek için HAR problemini bir görüntü sınıflandırma problemi olarak tanımlayan, Sinyal Görüntülerinde İnsan Aktivitesi Tanıma (HARSI) adlı bir yaklaşım önermektedir. Önerilen HARSI yaklaşımı, görüntü verilerinin işlenmesinde evrişimli sinir ağlarının (CNN) güçlü yönlerinden yararlanmak için akıllı sensör verilerini bazı görsel görüntülere dönüştürür. Gerçek dünya veri seti üzerinde gerçekleştirilen deneysel sonuçlar, önerilen HARSI modeli ile geleneksel makine öğrenme modellerine kıyasla önemli bir gelişme kaydedildiğini göstermiştir. Sonuçlar ayrıca, yöntemimizin sınıflandırma doğruluğu açısından son teknoloji yöntemlerden daha iyi performans gösterdiğini göstermiştir.

Anahtar kelimeler: İnsan aktivitesi tanıma, makine öğrenmesi, derin öğrenme, akıllı sensör sistemleri, açıklanabilir yapay zekâ.

CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE
INTRODUCTION

1.1 General

Human activity recognition (HAR) is the task of correctly identifying human activities (i.e., walking, eating, standing, and working) by analyzing smart sensor data. It is useful for understanding the behavioral human patterns in a smart sensor-based system (SSS). The HAR systems have been successfully used in various domains such as assisted living and healthcare (Hassan et al., 2020 and Kanjilal & Uysal, 2021), biometric user identification for security (Mekruksavanich & Jitpattanakul, 2021), well-being in smart homes (Mihoub, 2021), evaluating employee performances in smart factories for Industry 4.0 (Mohsen et al., 2021), body motion analysis in sports, and monitoring safety (falls, injuries, and collisions) in an environment (Madokoro et al., 2021).

Activity recognition is a significant indicator of participation, quality of life, and lifestyle. Human activities carry a lot of information about the context (i.e., a person's identity, personality, and mental state) and help systems to achieve context awareness. For example, patient activity recognition is critical in analyzing treatment progress and can provide context information for decision-making for better treatment and care. Similarly, rehabilitation specialists and therapists can remotely benefit from information on patient activities outside of a health center. Having detected the activities, a broad range of analyses (i.e., activities by age group, by gender, by location, by day of the week) can be performed to answer the questions of where and when users do which types of activities. It can help detect abnormalities in surveillance systems, therefore, preventing any unfavorable consequences. Using wearable sensors, HAR applications detect the actions of the users to provide them with intelligent personal assistance and recommendation. In the military, it is important to recognize the activities of the soldiers to provide feedback to their managers that assist them in real-time Consequently, there are numerous potential computing systems where recognizing human activities plays an important role.

The main issue with the current human activity recognition systems is they have been considered black-box systems, using non-understandable sensor data and providing predictions without being able to explain them. Without explainability and transparency, a HAR model is not trustworthy for making real-world decisions, especially the high-risk ones in smart sensor systems. This study aims to provide an interpretable, basic, reliable approach to HAR problems.

1.2 Main Contributions of the Thesis

The main contributions and novelty of this thesis can be listed as follows. (1) It proposes an approach, called Human Activity Recognition on Signal Images (HARSI), which converts the time series data to signal images and feeds them into a CNN for the classification task. (2) It provides an important attempt by improving human-level explainability for smart sensor data by using signal images in the field of HAR. (3) This study is also original in that it compares the performances of different CNN architectures on signal image data in terms of accuracy to determine the best one for HAR. (4) The proposed method outperforms both the classical machine learning methods (i.e., k-nearest neighbors, support vector machines, multi-layer perceptron, decision trees, AdaBoost, and random forest) and the state-of-the-art methods on the same dataset.

1.3 Purpose

The main purpose of this study is to build an effective deep-learning model that solves HAR problems. We developed the HARSI approach that accurately recognizes six human activities, including walking, jogging, standing, sitting, downstairs, and upstairs. The experimental results showed that a significant improvement was achieved by the proposed method compared to the traditional machine learning methods and the state-of-the-art methods in terms of recognition accuracy.

1.4 Organization of the Thesis

The thesis consists of seven chapters. The remainder of this thesis is structured as follows:

Chapter 2 explains recent studies on HAR that use a deep learning technique.

Chapter 3 describes deep learning and convolutional neural network.

Chapter 4 describes the proposed HARSI approach in detail.

Chapter 5 provides a brief description of the data and presents the experimental results.

Chapter 6 provides a discussion and compares the HARSI approach with other machine learning methods.

Chapter 7 gives conclusion remarks and opportunities for further research.

CHAPTER TWO
LITERATURE REVIEW

Research in the field of HAR is becoming increasingly important with the rapid development of smart sensor systems (Shalaby et al., 2022). Especially, HAR is of great significance in the Internet of Things (IoT) applications which include sensor and communication technologies. Different sensor types have been utilized in the HAR systems, including wearable sensors (Bijalwan et al., 2021, Ferrari et al., 2022, Manjarres et al., 2022, Bhat & Khan, 2021, Bozkurt, 2021), vision-based sensors/cameras (Khan et al., 2022, Khan et al., 2022, Tasnim et al., 2021), health sensors (Maitre et al., 2021), and environmental sensors (Madokoro et al., 2021, Hwang et al., 2021, Ronald et al., 2021). The ambient sensor-based HAR applications detect activities from the sensors that are installed at fixed locations (i.e., home, factory) or placed on a fixed object (i.e., fridge, door, toilet flush). In this study, we focused on wearable sensors since they provide many advantages like privacy protection, wide coverage area, and high robustness when modeling an activity classifier.

Wearable sensors are lightweight (few grams), small-size (few mm), easy to program, and low-cost. Using either a strap or an adhesive, they can be easily attached to many different body parts (i.e., arm, waist, shoulder, wrist, or leg) depending on the human activities being studied (Manjarres et al., 2022, Pei et al., 2021). The accelerometer (A) (Yen et al., 2021, Alawneh et al., 2021) is one of the widely used sensors to collect acceleration data related to human activity. Besides the accelerometer, gyroscope (G) and magnetometer (M) sensors are attached to the body in various ways for monitoring their actions at a particular point in time. The combination of the sensors (A, G, M) can also provide useful information when analyzed by machine learning methods to recognize human activities (Elsts & McConville, 2021). Most HAR systems (Ghate & C, 2021, Thakur & Biswas, 2021) have been currently developed by smartphones since phones contain several sensors (i.e., A, G, M) and they are practically wearable, useful, and easy-to-use computing devices with high computing power and storage.

Deep learning (DL) is one of the most exciting technologies that implement symmetry in computer science. In the literature, deep learning (DL) techniques have been proven to be powerful in activity classification. The recent studies on human activity recognition that used deep learning technology are given in Table 2.1. A variety of DL methods have been successfully used for HAR such as recurrent neural networks (RNN) (Alawneh et al., 2021, Ghate & C, 2021, Buffelli & Vandin, 2021, Alhersh et al., 2021) long short-term memory (LSTM) (Irfan et al., 2021, Al-Wesabi et al., 2021, Chen et al., 2020), autoencoder (AE) (Mihoub, 2021), deep neural network (DNN) (Hassan et al., 2020, Bijalwan et al., 2021, Bozkurt, 2021), and convolutional neural networks (CNN) (Hoai Thu & Han, 2021, Stuart & Manic, 2021).

Table 2.1 Summary of the recent human activity recognition studies that used deep learning techniques. (SG: Self-Generated)

<table>
<tr><th rowspan="2">Ref</th><th rowspan="2">Year</th><th colspan="5">Method</th><th rowspan="2">Description</th><th rowspan="2">Sensor Types</th><th rowspan="2">Data</th><th rowspan="2">Sensor Location</th><th rowspan="2">Number of Activities</th><th rowspan="2">Sensor-Data-Based</th><th rowspan="2">Signal-Image-Based</th><th rowspan="2">XAI</th></tr>
<tr><th>CNN</th><th>DNN</th><th>RNN</th><th>LSTM</th><th>AE</th></tr>
<tr><td>(Shalaby et al., 2022)</td><td>2022</td><td>√</td><td></td><td>√</td><td>√</td><td></td><td>Channel state information (CSI) based HAR</td><td>Wireless signal</td><td>CSI</td><td>Room</td><td>6</td><td>√</td><td>X</td><td>X</td></tr>
<tr><td>(Bijalwan et al., 2021)</td><td>2022</td><td>√</td><td>√</td><td></td><td>√</td><td></td><td>Gait pattern analysis</td><td>A, G, M</td><td>SG</td><td>Center of mass</td><td>7</td><td>√</td><td>X</td><td>X</td></tr>
<tr><td rowspan="3">(Ferrari et al., 2022)</td><td rowspan="3">2022</td><td rowspan="3">√</td><td rowspan="3"></td><td rowspan="3"></td><td rowspan="3"></td><td rowspan="3"></td><td rowspan="3">Personalization in HAR</td><td rowspan="3">A, G, M</td><td>UniMiB SHAR</td><td rowspan="3">Pocket</td><td>17</td><td rowspan="3">√</td><td rowspan="3">X</td><td rowspan="3">X</td></tr>
<tr><td>Motion Sense</td><td>6</td></tr>
<tr><td>MobiAct</td><td>15</td></tr>
</table>

Table 2.1 Continues

(Manjarres et al., 2022)	2022	√			√		HAR from piezoelectric-based kinetic energy signals	KEH transducers	KEH	Hand, Waist	5	√	X	X
(Bhat & Khan, 2021)	2022	√					Feature extraction-based approach	A	WISDM	Pocket	6	√	X	X
(Bozkurt, 2021)	2022	√	√		√		Comparative study on classifying human activities	A, G	UCI-HAR	Waist	6	√	X	X
(Khan et al., 2022)	2022	√		√			Gesture recognition in videos	Camera	SG	Room	4	√	X	X
(Hassan et al., 2020)	2021		√				HAR from highly sparse body sensor data	A, RFID	Room set1 Room set2	Chest	4	√	X	X
(Kanjilal & Uysal, 2021)	2021	√		√	√		Feature extraction-based approach	A, G, M	UniMiB-SHAR	Pocket	17	√	X	X

Table 2.1 Continues

(Mekruksavanich & Jitpattanakul, 2021)	2021	√			√		Biometric user identification	A, G	UCI-HAR USC-HAD	Waist	6 12	√	X	X
(Mihoub, 2021)	2021			√	√	√	HAR in smart homes	Env. sensors	Orange4Home	Room	24	√	X	X
(Mohsen et al., 2021)	2021	√			√		Industry 4.0-oriented approach	A	WISDM	Pocket	6	√	X	X
(Madokoro et al., 2021)	2021	√			√		Human pose and motion estimation	Camera Env. sensors	SG	Room	5 8	√	X	X
(Khan et al., 2022)	2021	√			√		Hybrid deep learning-based model	Motion Kinect sensor	SG	Room	12	√	X	X
(Tasnim et al., 2021)	2021	√					HAR using skeleton datasets	Camera	UTD-MHAD MSR-Action3D	Room	27 20	√	X	X
(Maitre et al., 2021)	2021	√					Feature fusion-based approach	A, G, M	MHEALTH	Ankle, Arm, Chest	12	√	X	X

Table 2.1 Continues

(Hwang et al., 2021)	2021	√			√		Causality feature extraction based approach	Env. sensors	Aruba	Room	10	√	X	X
									Milan		15			
									Cairo		13			
(Ronald et al., 2021)	2021	√			√		HAR based on the Inception-ResNet model	A, G	UCI-HAR	Waist	6	√	X	X
								Env. and body sensors	Opportunity	Room, Body	18			
								A	Daphnet	Legs, Hip	2			
								A, G, M	PAMAP2	Chest, Ankle	18			
(Pei et al., 2021)	2021				√	√	Multiple domain DL framework	A, G, M	SG	Head, Wrist, Leg	12	√	X	X
(Yen et al., 2021)	2021	√					Feature fusion-based approach	A, G	SG	Waist	6	√	X	X
									UCI-HAR					
(Irfan et al., 2021)	2021	√			√		Recognizing transitional activities	A, G	HAPT	Waist	12	√	X	X
									HAD		5			

Table 2.1 Continues

(Alawneh et al., 2021)	2021			√	√		HAR using time series data	A	UniMiB SHAR	Pocket	17	√	X	X
								A	WISDM	Pocket	6			
								A	UCI-HAR	Waist	6			
(Elsts & McConville, 2021)	2021	√					HAR on microcontrollers	A, G, M	PAMAP2	Hand, Chest, Ankle	12	√	X	X
(Ghate & C, 2021)	2021	√		√	√		Hybrid deep learning-based approach	A, G	UCI-HAR	Waist	6	√	X	X
								A	WISDM	Pocket				
(Thakur & Biswas, 2021)	2021	√					Feature fusion-based approach	A, G	SG	Pocket	6	√	X	X
									UCI-HAR	Waist				
(Buffelli & Vandin, 2021)	2021	√		√	√		Attention-based mechanism	A, G	HHAR	Hand, Chest, Ankle	6	√	X	X
								A, G, M	PAMAP2		12			
								A, G	USC-HAD		12			
(Alhersh et al., 2021)	2021	√		√			HAR using multimodal sensors	Multi-modal	CMU-MMAC	Room	11	√	X	X
(Chen et al., 2020)	2021				√		Multimodal complex HAR	A, G, M	Lifelog	Pocket, Wrist, Chest	9	√	X	X
									PAMAP2	Wrist, Arm, Chest,	18			

Table 2.1 Continues

(Hoai Thu & Han, 2021)	2021	√			√		Hierarchical hybrid deep learning-based approach	A, G	UCI-HAPT MobiAct	Waist Pocket	12 11	√	X	X
(Stuart & Manic, 2021)	2021	√					Resource-constrained HAR	EMG sensors	Myo-TL Db5	Elbow, Wrist	9 18	√	X	X
Our Approach		√					**Human activity recognition on signal images (HARSI)**	A	WISDM	Pocket	6		√	√

In the literature, most HAR studies (Bhat & Khan, 2021, Bozkurt, 2021, Yen et al., 2021, Khan et al., 2022) focused on the identification of daily living activities such as standing, sitting, and walking. However, some previous studies try to detect more specific types of activities such as cheating activity in an exam (Khan et al., 2022), rope jumping (Elsts & McConville, 2021), shopping (Chen et al., 2020), housekeeping (Hwang et al., 2021), and sports activities (i.e., basketball, bowling, boxing, and tennis). Besides these high-level activities, some works (Kanjilal & Uysal, 2021, Irfan et al., 2021) have been focused on the transitions between the activities such as stand-to-sit or sit-to-stand. In this study, we built a CNN model to recognize six different activities, including walking, jogging, standing, sitting, downstairs, and upstairs.

Typically, a HAR framework contains the following stages: data collection, data pre-processing, feature extraction, feature selection, training, performance evaluation, classification, and decision-making. Using raw sensor data directly in machine learning is usually not practical since it doesn’t carry sufficient information to distinguish different human activities. In other words, only one particular value at a specific time instant of an action doesn't carry enough information to describe the activity itself. For this reason, the standard HAR studies involve the feature extraction phase to transform time-series data into samples that summarize the data over a fixed period. In general, frequency-domain features and time-domain features are extracted from the raw sensor data, such as max, min, median, mean, peak-to-peak value, standard deviation, the number of zero crossings, skewness, kurtosis, and signal entropy [Anna Ferrari]. In the literature, some studies (Maitre et al., 2021, Yen et al., 2021, Thakur & Biswas, 2021) mainly concentrated on this issue since it plays an important role in the final system performance.

HAR models can be considered in two categories: per-subject (personalized) and cross-subject (generalized) models. In per-subject models, both the training and testing data are all from the same individual since each person has unique characteristics of movements, such as speed corresponding to its physical properties (i.e., age, gender, height, and weight) and habits. On the other hand, in cross-subject models, the training and testing data come from all the persons. They provide many advantages such as working on a large amount of data to build a robust classifier and dealing with a single classifier, instead of multiple ones.

Our study differs from the studies aforementioned here in several aspects. In our study, we do not use the features that were extracted from sensor data like most HAR systems; rather, we transfer time-series data into signal images that reflect the properties of activities. Here, we present a detailed analysis of the performances of different CNN architectures on human activity signal image data for the first time. In addition, in this study, we provide human-level explainability for smart sensor data in the field of HAR. We aim to provide an explainable artificial intelligence (XAI) approach that can give human-understandable information and prediction. To the best of our knowledge, both concepts together (signal image-based HAR and XAI) have not been studied so far.

CHAPTER THREE
BACKGROUND INFORMATION

3.1 Deep Learning

Deep learning is a subset of machine learning that uses neural networks consisting of multiple hidden layers for prediction tasks. The main learning types are supervised learning, semi-supervised learning, and unsupervised learning.

Deep learning architectures like convolutional neural networks have been applied to many research areas such as computer vision, human activity recognition, natural language processing, drug production, analysis of images in medical applications, and environmental science.

"Deep" adjective shows using layers more than one in a neural network. Deep learning is a modern variation that is related to an unlimited number of network layers of bounded size (Wikimedia Foundation, 2022)

3.2 Convolutional Neural Network (CNN)

CNN is a deep artificial neural network that generally contains more than one different neural network layer. Each layer contains many neurons. CNN has many layers in it such as the pooling layer, convolutional layer, and fully-connected layer. The complete design of a CNN is shown in Figure 3.1. In the figure, there are 3 pooling layers, 1 input layer, 1 output layer, a fully connected layer, and 3 convolutional layers. Data is taken by the input layer, and it is preprocessed in this layer. Features are extracted at the convolutional layer. The pooling layer decreases the number of features. At the intermediate level, convolutional layers work together with the pooling layers. The fully connected layer is responsible for summarizing all results from each layer, and it gives a final result. The last layer, namely the output layer, produces algorithm results which are extracted by the connected layer and its outputs. (Wan et al., 2019).

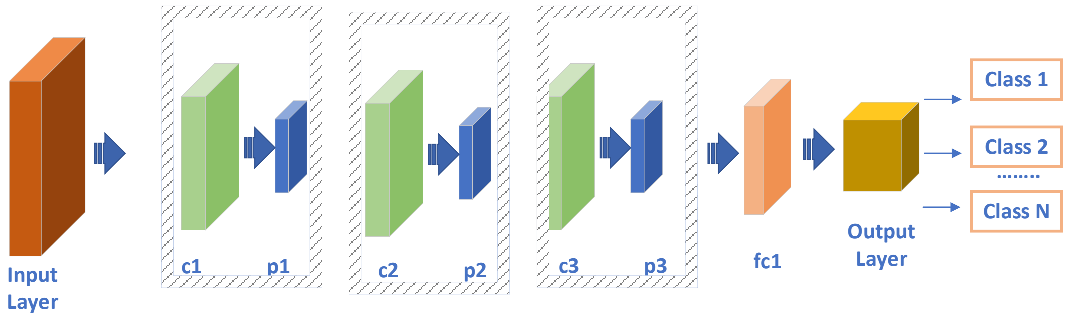

Figure 3.1. Convolutional Neural Network model

CHAPTER FOUR
MATERIALS AND METHODS

4.1 General Description

Human activity recognition is a problem of classifying data obtained from sensors into well-defined actions performed by humans, e.g., walking, jogging, and sitting. Recognition of activities is a challenging time-dependent task since there is no single and precise way or formula to define specific movements. Machine learning methods have played a major role in the analysis of sensor data for providing real-time feedback in HAR applications.

In a HAR system, an accelerometer sensor embedded in an IoT device is placed at a specific location on the human body and is synchronized to emit data in an IoT environment. The accelerometer measures three-dimensional acceleration referenced with the earth's gravity during dynamic states. The sensor generates a signal along the *x*-axis, *y*-axis, and *z*-axis at a time step $t \in \{1,2,\ldots, T\}$. Accelerometer measurements are collected in a time interval for a person. For this sensor, the sensing data along time can be represented by a multi-dimensional time series $S = [\ s_1, s_2, \ldots, s_t, \ldots]$ where s_t is the sensing signal of the sensor placed at a body location at time *t*. Each signal record has six attributes {datetime, sensorID, x-axis, y-axis, z-axis, activity}.

A single measurement is not enough to classify an activity because of the time-dependent nature of activities. To deal with this problem, *S* is segmented into multiple frames, called windows, and then, each frame is mapped into a predefined activity label. In our study, temporal segmentation through the sliding window technique is necessary to define the boundaries of signal images. Each image is annotated by an activity from a label set, which is denoted by $a_i \in \{A_1, A_2, \ldots, A_m\}$, where m is the number of potential activities to be recognized. For instance, the class labels can be as follows: A_1=walking, A_2=sitting, and A_3=stairs.

The activity recognition problem can be described as follows. Given a recorded signal time series *S*, the task is to detect an activity (e.g., walking) that infers human behavior in a period.

In the proposed approach, time series data is converted to images by drawing three lines according to the recorded *x*-*y*-*z* values, and then, the images are fed into a CNN for the image classification task.

In a traditional HAR system, features used for activity classification are commonly extracted from time-series sensor data by using statistical methods such as max, min, mean, peak-to-peak value, standard deviation, the number of zero crossings, skewness, kurtosis, and entropy. However, this input data cannot be readable and interpretable by humans, as can be seen in Figure 4.1. To overcome this problem, we propose an approach that provides human-level explainability for smart sensor data. For instance, when the numeric feature vector [339,27,0.3,0.4,0.08,0.06,0.05,0.07, ...] is seen by a human, it cannot be directly associated with the "walking" activity since it lacks visual representation.

339,27,0.35,0.4,0.08,0.06,0.05,0.07,0,0,0,0,0,0.06,0.05,0.08,0.09,0.06,0.13,0.13,0.24,0,0.09,0.1,0.02,0.12,0.2,0.14,0.05,0.11,0.1,0,0,10,-0.23,1193.75,561.76,2625,2.39,4.45,2.72,3.2,5.13,3.2,11.14,Walking

1,6,0,0,0,0,0,0.02,0.04,0.1,0.06,0.01,0.01,0.02,0.07,0.08,0.02,0.02,0,0.01,0,0,0.02,0.01,0.01,0,0.05,0.07,0.02,0.01,0.03,0.01,0,0.23,-0.02,625,650,900,2.16,0.5,0.68,10.94,3.29,10.94,2.43,Jogging

3,15,0.1,0.1,0.1,0.1,0.09,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.09,0.1,0.09,0.1,0.1,0.1,0.09,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.08,0.12,0,8.92,-0.21,300,1683.33,2200,3.49,3.34,1.78,4.41,4.1,4.41,10.19,Upstairs

2,13,0.1,0.11,0.1,0.14,0.1,0.1,0.1,0.08,0.1,0,0.1,0.09,0.1,0.1,0.12,0.13,0.09,0.12,0.09,0,0.09,0.13,0.09,0.1,0.11,0.1,0.1,0.12,0.1,0,0,9.56,2.02,1000,850,3450,1.27,1.98,1.55,1.62,2.75,1.62,10.1,Downstairs

428,27,0.21,0.23,0.2,0,0.12,0.07,0,0.06,0.02,0.11,0.1,0.11,0,0,0.21,0,0,0.19,0.14,0.26,0.07,0.03,0,0.02,0.04,0,0.04,0.11,0.17,0.52,0,9.32,1.23,510.71,203.19,275,3.12,0.09,0.1,3.13,0.21,3.13,9.91,Sitting

432,35,0.2,0.42,0.12,0.1,0,0.05,0,0.03,0.01,0.05,0.05,0.07,0,0.08,0,0,0.29,0,0,0.47,0.09,0.12,0.25,0,0,0.25,0,0.15,0,0.16,0,9.36,2.82,300,260.81,285.29,2.39,0.05,0.05,2.4,0.2,2.4,10.07,Standing

Figure 4.1 Sensor data that cannot readable and interpretable by humans.

Rather than extracting features from time-series sensor data, we assemble *x*-axis, *y*-axis, and *z*-axis accelerometer signal sequences into an image to enable CNNs to learn the optimal features automatically from the signal image for the human activity classification task. In other words, we propose an approach, called HARSI, which transforms numerical sensor data into image format data and builds a CNN model that enables human activity recognition on these signal images. The main purpose of our study is to provide an interpretable and robust approach to HAR problems.

To provide human-level explainability for smart sensor data, we propose an approach that visualizes the data with charts, as can be seen in Figure 4.2. Generating signal images makes data understandable for humans. Each chart can be easily associated with activities by humans. For example, Figure 4.2 corresponds to the "jogging" activity since both the amplitude and frequency of the signal are high as a result of the high velocity and displacement of the person on the ground.

Figure 4.2 shows sample signal images that include the *x*, *y*, and *z* axes values of an accelerometer sensor (A_x, A_y, and A_z) over time for each activity. The activity images can be easily understandable, interpretable, and explainable by humans since each one has different characteristics. For example, walking is more periodic than standing. Jogging requires higher effort and power than walking since it requires more intense muscle contractions. It can also be seen that very low *x*-*y*-*z*-values are observed for sitting. Moreover, a human may require different relaxation times when going upstairs, and sometimes, he/she stops to have a rest, moves slowly, and spends more time because he/she feels tired. The acceleration curve of going downstairs is similar to going upstairs, but the cycle of motion is shorter. These observations are also consistent with the energy harvesting from human activities. Sitting and standing activities are non-periodic since they are stable and straightforward postures, while other activities are mainly repetitive and quasi-periodic.

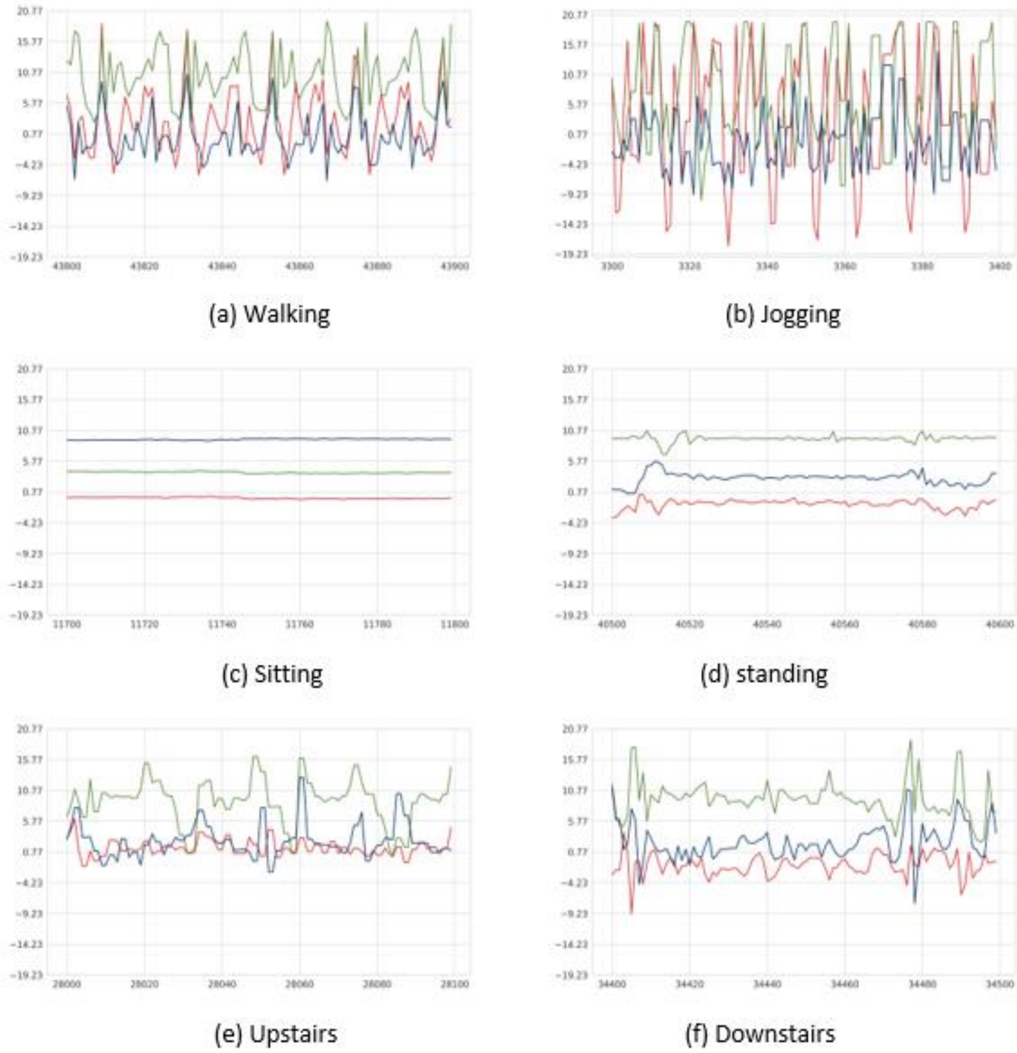

Figure 4.2 Explainable sensor data for each human activity.

Figure 4.3 shows a general overview of the proposed HARSI approach. The approach mainly includes the following stages: data collection, data transformation, training, testing, and classification. The data collection stage comprises obtaining raw signal values from an accelerometer sensor embedded in a smartphone when performing human actions. After that, the collected raw data is transferred to a server through communication technology such as Bluetooth or WiFi. In the data transformation stage, the time-series signal data is divided into fixed-size segments, called windows, by using a sliding-window method. After that, an image is generated for each window by drawing lines on sample values. Here, each signal image corresponds to a single activity such as standing, sitting, or walking. In the training stage, each signal image is fed to CNN as an input vector, and then CNN learns different features of the image through different layers. In the classification stage, the CNN model gives insight according to the features of the signal image. In other words, the CNN model predicts a given input image according to the class probabilities of activities, such as standing, sitting, jogging, walking, upstairs, and downstairs. In the testing stage, the performance of the CNN model is assessed by using a test set to evaluate how well it recognizes human activities. If the prediction accuracy of the CNN model is an acceptable level (i.e., >80%), it can be further used to recognize real-time human activities. Afterward, the final prediction can be considered in a decision support stage to guide the decision-maker. Each stage in the workflow involves several research questions to be addressed. Furthermore, since the environment of HAR systems is dynamic and ever-evolving, it is required to update the model periodically by following the same stages to achieve high accuracy consistently.

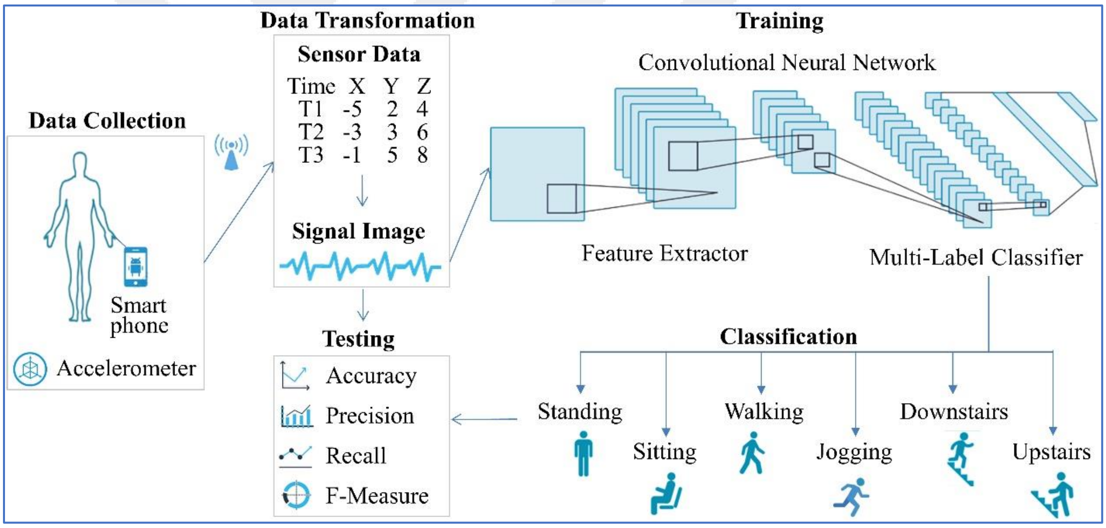

Figure 4.3 The general workflow of the proposed HARSI approach.

As seen in Figure 4.3, the CNN contains an input layer, multiple hidden layers, and an output layer. Signal image data are processed layer-by-layer, where the output of each layer becomes the input for the next layer. Each layer contains multiple units, which are denoted by U_i^l to indicate the ith unit in layer l. The hidden layers are composed of convolutional, pooling, and fully connected layers.

Convolutional layer (CL): They are used as feature extractors to automatically obtain high-level representations of input images. Formally, a feature map is extracted using a convolution procedure as follows:

$$F_j^{l+1} = \alpha\left(\sum_{i=1}^{|F^l|} K_{j,i}^l F_j^l + b_j^l\right) \tag{4.1}$$

where F_j^l denotes the jth feature map in layer l, $|F^l|$ is the number of feature maps in layer l, $\alpha()$ is an activation function, b_j^l is a bias vector, and $K_{(j,i)}^l$ represents the kernel applied on feature map i in layer l to obtain jth feature map in layer $(l + 1)$.

Pooling layer (PL): Pooling layers are used to reduce dimensionality, as well as the number of parameters. A PL is usually inserted between successive CLs in a CNN architecture. Formally, max pooling is given by:

$$v_i^{l+1} = \max_{1 \le k \le r} (v_{i+k}^l) \tag{4.2}$$

where r is the pooling size and v_i^l refers to the value of the ith unit in layer l.

Fully connected layer (FCL): After multiple CLs and PLs, the classification process is handled in a fully connected layer, which produces an output vector as given in Equation (3)

$$\mathbf{z}^{l+1} = v_i^l \mathbf{w} \tag{4.3}$$

where w represents a weight vector and vector z includes non-normalized log probabilities. The output of the FCL is fed into a softmax classifier, which predicts the activity label as follows:

$$Softmax(\mathbf{z}_i) = \mathrm{P}(o = a \mid \mathbf{z}_i) = \frac{e^{z_i}}{\sum_{j=1}^{m} e^{z_j}} \tag{4.4}$$

where a is an activity label, o denotes the output of the classification model, z_j represents the jth element of log probability vector z, and m is the number of class labels. The predicted activity label (al) for a given image is assigned to the one with the highest probability as given in Equation (5).

$$al \leftarrow \mathrm{argmax}_{a=1}^{m} \mathrm{P}(o = a \mid image) \tag{4.5}$$

4.2 Formal Definition

Let the raw dataset D be a set of instances collected by an accelerometer sensor. Each instance in dataset D includes a set of pairs of x-y-z axis values and the corresponding activity label, which is denoted by $D = \{(x_1, y_1, z_1, a_1), (x_2, y_2, z_2, a_2). (x_n, y_n, z_n, a_n)\}$, where n is the number of instances. In other words, a_i is the activity (class label) belonging to the axes values of the sensor (x_i, y_i, z_i). The output attribute $O = \{a_1, a_2, ... , a_n\}$ has m different human activities, which is denoted by $a_i \in \{A_1, A_2, ... , A_m\}$ for $i = 1, 2, \ldots , n$. For example, in a four-activity classification (sitting, standing, stairs, walking), the class labels of the instances are A_1=sitting, A_2=standing, A_3=stairs, and A_4=walking.

In the proposed HARSI approach, the raw sensor data is transformed into signal images using a sliding window method. In this process, a large time-series data is split into fixed-sized chunks, referred to as windows, and a single activity label is assigned to all the samples within the window. A window is defined as a set of consecutive sequences such that $S = \{s_r, s_{r+1}, ..., s_{r+w-1}\}$, where w refers to the window size and r corresponds to an arbitrary position, such as $1 \leq r \leq n\text{-}w+1$, where n is the data size. An image is generated for each window by drawing lines on sample values. Each signal image is labeled with the corresponding activity a. A sliding-window method can be performed in either a non-overlapping or overlapping way. A non-overlapping method indicates that the values in one image do not intersect with the values of the other successive image, i.e., $w_1 \cap w2 = \emptyset$. On the other hand, an overlapping method is defined by a particular percentage, which indicates how many samples from the previous image are repeated in the current image, i.e., $w_1 \cap w2 \neq \emptyset$. In this study, we prefer to use the non-overlapping image technique to prevent information duplication.

Definition 1 (window). A window is defined as a set of consecutive sensor measurements obtained within a time interval such that $w = \{s_r, s_{r+1}, ..., s_{r+q-1}\}$, where q refers to the window size and r corresponds to an arbitrary position, such as $1 \leq r \leq n\text{-}q\text{+}1$, where n is the data size.

After generating windows, a single activity label $a_i \in \{A_1, A_2, …. , A_m\}$ is assigned to each window $\{(w_1, a_i), (w_2, a_i),…, (w_{n/q}, a_i)\}$ such that all the samples within the window belong to the respective class. After that, an image is generated for each window by drawing lines on sample values. Each signal image is labeled with the corresponding activity a.

Definition 2 (activity). An activity is a human movement characterized by a body action or posture, e.g., walking. An activity label $a_i \in \{A_1, A_2, …. , A_m\}$ is associated with an image that is generated from a window with fixed length (q) by segmenting the raw sensor data D, where m is the number of potential activities to be recognized.

The problem studied in this study is to detect a corresponding activity implicated in a certain temporal sequence based on the classification. In other words, the aim of HAR is to build a model M(image, •) to infer the correct activity label for a given image, where • denotes all the parameters to be learned during the training process.

Definition 3 (activity recognition task). Given a set of training images with their corresponding activity labels and a query image, the aim is to find a mapping function f : image → activity that correctly infers the human behavior for the query image. The predicted activity label should be as similar as possible to the actual class label. Therefore, the task is to build a classification model M by minimizing total loss $L(M)$.

A crucial factor in a sliding window method is to select a suitable window size to achieve high recognition accuracy since the ideal window size varies in accordance with the characteristics of the signals being processed. In general, the small window sizes can be useful to detect faster-changing activities better; however, using short windows may lead to misclassification because some vital information about a complex activity may not be captured by multiple windows. On the other hand, large windows provide to detect complex activities and semi-complex activities. However, a large window generates a signal image that belongs to more than one human activity, and this leads to a decrease in recognition accuracy. Considering this tradeoff, researchers usually determine the optimal window size by trying empirical values and assessing classification accuracy. In this study, the size of each window was set to 100 samples since the dataset was collected at a rate of 20 samples per second, and it is a sufficient sampling value to make a reasonable prediction for human activity.

Figure 4.4 shows the pseudocode of the proposed HARSI method. In the algorithm, first, a sliding window technique is used to split accelerometers sensor data (*D*) into windows with size w. In other words, it segments data streams into windows of equal length. For a dataset with n samples, the algorithm generates *n/w* windows, where each one ranges between *i*q* and *i*q+q-1* for *i = 0,1,...,n/q*.. After that, an image is generated for each window by drawing lines on sample values. Here, a small window is shifted along the continuous data stream converting contiguous portions of sensor readings into images. Each image is labeled with the corresponding activity. A CNN classifier *M* is then trained on the signal image dataset. In the final step, the activity (class label) of each unseen image in the test set *T* is predicted by using the classifier.

Two-stage detectors such as Regions with Convolutional Neural Networks (R-CNN) (Girshick, Donahue, Darrell, & Malik, 2014), Faster R-CNN (Ren, He, Girshick, & Sun, 2017), Feature Pyramid Network (FPN) (Lin et al., 2017) classify objects in two phases. In the first phase, candidate regions are generated from the input image and after that, classes and bounding boxes are created using these regions.

Algorithm Human Activity Recognition on Signal Images (HARSI)

Inputs: $D = \{(x_1, y_1, z_1, a_1), (x_2, y_2, z_2, a_2), \ldots, (x_n, y_n, z_n, a_n)\}$
q: window size
T: Test set

Output: $O = \{o_1, o_2, \ldots, o_t\}$ a set of outputs for test images

Begin:

```
    for i = 0 to n/q do
      W = Ø
      for j = i *q to i * q + q - 1 do
        W = W U (x_j, y_j, z_j,)
        activity = a_j
      end for
      image = ConvertToImage(W)
      I = I U <image, activity>
    end for
    M = CNN(I)
    foreach image i in T do
      o = Classify(M, i)
      O = O U o
    end foreach
    Return O
```

End

Figure 4.4 HARSI approach algorithm

CHAPTER FIVE

EXPERIMENTAL STUDIES AND RESULTS

This section presents a detailed study that had been carried out to evaluate the performance of the proposed HARSI method. The effectiveness of the method was demonstrated on a real-world dataset by using different CNN architectures, including AlexNet, ResNet, SqueezeNet, DenseNet, and VGG. The parameter settings and the number of parameters for each CNN architecture are given in Table 2.1. The structures of CNNs are different from each other in several aspects such as the number of parameters and the number of layers. For instance, ResNet34 consists of a residual network with 34 layers. Compared to ResNet50, the DenseNet121 has more layers, whilst VGG19 has fewer layers. Some parameter settings are common in all models, i.e., rectified linear unit (ReLU) was used as an activation function and the output probability was calculated by Softmax. As the nature of the models used, batch layers predate each ReLU. In all models, Adam and Cross Entropy were used as the optimizer and loss function, respectively. These techniques have lately gained popularity and performed promising outcomes for deep learning applications.

In this study, optimal learning rate parameters were determined for each model separately to speed up the training process, adapt itself to the problem, and strengthen the generalization ability of the classifiers. While a low learning rate slows the convergence of the training process, a high learning rate can cause an unpleasant divergence in performance. Therefore, a suitable learning rate is vital for obtaining a satisfactory performance, however, finding an appropriate learning rate is both laborious and hard to decide. To solve this problem, we used the lr_find() method in Fast.AI, which is a deep-learning library built on top of PyTorch. This method works on the principle of using a very low learning rate initially to train a mini-batch and calculate the loss. In the next step, the method trains the next mini-batch with a small-scale higher learning rate than the previous one until it finds a learning rate where the model diverges. The optimal learning rate values determined for each model are listed in Table 5.1.

Table 5.1 Parameter settings and the number of parameters.

Model	Learn Rate	Activation Function	Optimizer	Loss Function	Total parameters	Total trainable parameters	Total Non-trainable parameters
HARSI ResNet34	2e-3	RELU (Rectified Linear Unity)	Adaptive Moment Estimation	Cross Entropy	21,815,104	547,456	21,267,648
HARSI ResNet50	6e-4				25,617,472	2,162,560	23,454,912
HARSI ResNet101	1e-3				44,609,600	2,214,784	42,394,816
HARSI AlexNet	2e-3				2,736,960	267,264	2,469,696
HARSI DenseNet121	4e-4				8,010,624	1,140,416	6,870,208
HARSI SqueezeNet_v1.0	2e-3				1,265,856	530,432	735,424
HARSI SqueezeNet_v1.1	1e-3				1,252,928	530,432	722,496
HARSI - VGG16	1e-3				15,253,568	538,880	14,714,688
HARSI - VGG19	2e-3				20,565,824	541,440	20,024,384

One of the main differences between the CNN models is the number of parameters, which can reflect the computational complexity of the model. It may be noted that as the number of total parameters increases, the time required for training usually increases. There are two categories of parameters, one is trainable parameters (i.e., weights of connections between layers) that are continuously updated to reduce the loss, and the other one is non-trainable parameters (i.e., biases) that are not optimized during the training process. For instance, in VGG19 architecture, the number of trainable parameters is 541,440, while the number of non-trainable parameters is approximately 20 million. As seen in Table 5.1, SqueezeNet has the smallest total number of parameters than others, whilst ResNet101 has the largest. The architectures have approximately 2, 15, and 25 million parameters for AlexNet, VGG16, and ResNet50, respectively. These values may be associated with computational complexity, where the higher the number of parameters, the greater the computational load during the training process.

The method was implemented in Python by using the PyTorch framework and various libraries such as Fastai, NumPy, Pandas, Scikit-Learn, Matplotlib, and Seaborn. In this study, the CNN models were trained on a computer equipped with an Nvidia GTX 1060 graphics card using the Cuda toolkit to make use of the GPU computational capability and reduce implementation time through a rapid and simple design.

5.1 Dataset Description

In order to show the effectiveness of the proposed approach, the WISDM (Wireless Sensor Data Mining) dataset (Kwapisz et al., 2011) was used in the experiments. It was released by the Laboratory at Fordham University in the United States. This dataset is one of the important and popular large-scale benchmark datasets in the field of HAR. It has been used in many studies so it is suitable for making comparisons with previous works. (Baysari et al., 2022, Cengiz et al., 2022) The dataset is appropriate for detecting activities. Before collecting the data, the researchers obtained approval from the University Board since it was research on human subjects and involved some risks, i.e., the subject could fall down while jogging. The data collection process was fully monitored and guided by researchers in the laboratory environment to ensure the quality of the data. The dataset contains routine motion patterns with a significant number of processable movement samples. The dataset has 1,098,207 samples that were collected from 36 different participants while performing six activities. The percentages of each activity in the dataset are as follows: Jogging 31.2%, Downstairs 9.1%, Walking 38.6%, Upstairs 11.2%, Standing 4.4%, and Sitting 5.5%. While collecting sensor data, the participants are requested to carry a smartphone in their front pockets. With this experiment setup, accelerometer data were retrieved at each 50 ms, which means 20 samples per second from a smartphone while participants are acting activities. The raw dataset consists of x-, y-, and z-axis values obtained by an accelerometer sensor.

In this study, image representations are first generated from the raw x, y, and z values in the accelerometer sensor data. In other words, we converted the time series data into signal images by drawing three lines on sample x-y-z values. Here, a small window was shifted along the continuous data stream converting contiguous portions of sensor readings into images. In the generated images, x, y, z drawing lines are represented with different colors; red, green, and blue, respectively. While generating charts from sensor data values, attention was paid to choosing a fixed range for all the graphs. Accordingly, the max and min values were found by searching the x, y, z axes values in the dataset. The vertical range of the graph in each image approximately lies between [20, -20]. The horizontal range of the graph was set to 100 samples since the dataset was collected at a rate of 20 samples per second, and therefore it is a sufficient range to make a reasonable prediction for activity. Each image is labeled with the corresponding activity, such as standing, sitting, or walking. To create a balanced dataset, 400 images were generated for each activity; therefore, in total, 2400 images were generated. Figure 4.2 shows sample signal images for each activity.

By transforming numerical sensor data into image data, we aim to improve both explainability and recognition accuracy. The generated images provide human-level explainability for smart sensor data. Since each image reflects the properties of activities, they can be easily interpretable by humans. Rather than solving a time-series data classification problem, we define the HAR problem as an image classification problem. In this way, we provide an interpretable and robust approach to HAR problems.

5.2 Experimental Studies

In this study, we split the dataset into an 80% training set and 20% test set. The test set contains 100 images from each category, therefore, it includes 600 images in total. Four different metrics were used to evaluate the performance of each CNN architecture: accuracy, recall, precision, and f-measure. Accuracy is the fraction of correct predictions of the model to total prediction. Equation (1) shows how the accuracy rate is calculated,

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{5.1}$$

where TP is true-positive, FP is false-positive, TN is true-negative, and FN is false-negative. Precision provides how precise the model is out of the samples predicted positive, and how many of them are actually positive. Recall indicates how many of the actual positives the model capture through classifying it as positive. F-measure offers a single score that balances both the concerns of recall and precision values. Equations (2), (3), and (4) show the calculations of precision, recall, and f-measure values, respectively.

$$Precision = \frac{TP}{TP + FP} \tag{5.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{5.3}$$

$$Fmeasure = \frac{2 \times precision \times recall}{precision + recall} \tag{5.4}$$

5.3 Experimental Results

The effectiveness of the proposed approach (HARSI) was demonstrated on a real-world dataset by using different CNN architectures, including Alex Network (AlexNet), Residual Network (ResNet), Visual Geometry Group (VGG) Network, SqueezeNet, and Dense Convolutional Network (DenseNet). These CNN architectures were selected because of their popularity, high robustness, proven efficiency, and ability in image classification.

Table 5.2 shows the performance of the proposed HARSI method on different CNN architectures for the same dataset. Based on the accuracy rates, it is possible to say that all the CNN models have good classification ability. However, VGG19 is the most successful model among them with a 98% of success rate. Following this, ResNet34 has also a high accuracy rate (97.33%) in distinguishing human activities. This success is the result of the strengths of CNNs in classifying images. CNNs are capable of extracting key features directly and effectively from images, learning useful information layer-by-layer, and successfully classifying them in different classes.

Table 5.2 The performance of the proposed HARSI method on different CNN architectures.

Model	**Accuracy (%)**	**Precision**	**Recall**	**F-Measure**
HARSI - ResNet34	97.33	0.97433	0.97333	0.97326
HARSI - ResNet50	96.00	0.96054	0.96000	0.96006
HARSI - ResNet101	97.17	0.97194	0.97166	0.97161
HARSI - AlexNet	89.17	0.89018	0.89166	0.89077
HARSI - DenseNet121	96.67	0.96669	0.96666	0.96667
HARSI - SqueezeNet_v1.0	89.67	0.90025	0.89666	0.89725
HARSI - SqueezeNet_v1.1	93.00	0.92915	0.93000	0.92937
HARSI - VGG16	96.83	0.96871	0.96833	0.96826
HARSI - VGG19	**98.00**	**0.97999**	**0.98000**	**0.97999**

In addition to accuracy, we also evaluated the performance of the proposed HARSI approach on different CNN architectures in terms of recall, precision, and f-measure metrics. The values of these metrics are ranged between 0 and 1, where 1 is the best value. As can be seen in Table 5.2, the recall value obtained by the VGG19 model is closer to 1 than the others. This means that the VGG19 model often tends to give better predictions than the rest. As can be observed, the VGG16 model also outperformed the others in terms of precision and f-measure, as well.

In Figure 5.1, the loss values in both the training and validation processes are shown. While the vertical axis indicates the loss value, the horizontal axis represents the number of batches processed. In initial batches, the training loss is higher than the validation loss. As can be seen, both the training and validation losses reduce with the increase of the batches. The training loss and validation loss converged after approximately 200 batches were processed. When the minimum validation loss was obtained, the training process was stopped to avoid overfitting.

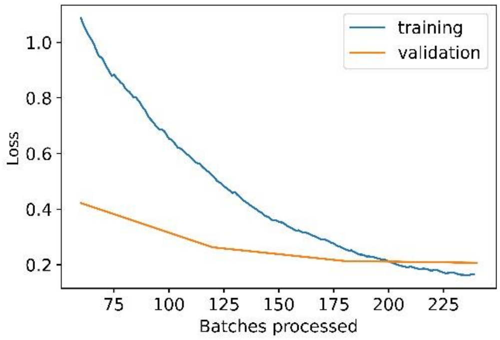

Figure 5.1 Loss values in the training and validation processes.

Figure 5.2 presents the confusion matrix to show the predictive performance of the proposed HARSI method on each human activity separately. According to the matrix, it is possible to say that the model usually had no difficulty in distinguishing human activities. For example, 98 out of 100 walking activities were predicted correctly; however, only 2 of the walking activities were misclassified by the classifier. Although each activity was recognized with a high accuracy rate, downstairs and upstairs activities were slightly confused with each other. It can be concluded from the confusion matrix that the best accuracies were achieved on the sitting, standing, and jogging activities.

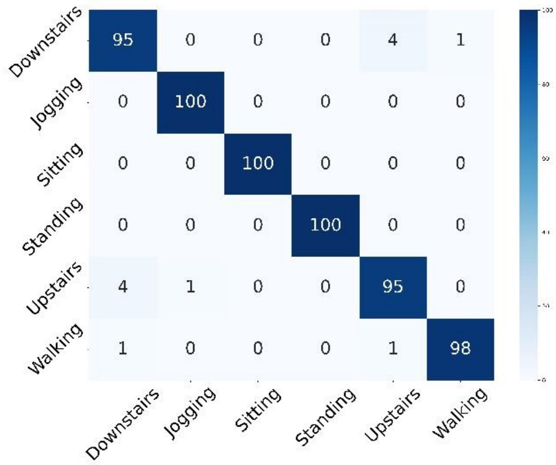

Figure 5.2 Confusion matrix obtained by the proposed HARSI method.

Figure 5.3 shows the execution times of the proposed HARSI method on different CNN architectures in minutes. Although all the times are close to each other, AlexNet and SqueezeNet are the fastest ones among their counterparts. They are followed by the ResNet34 model (1.11 minutes). The VGG models are also efficient in terms of training time (1.19 and 1.24 minutes). The DenseNet model may take a longer time (1.28 minutes), especially handling large image datasets. This is probably because of the higher number of layers in its architecture. Similarly, the required time for training ResNet101 is higher than others since it has a higher number of parameters to be assessed. The size and resolution of the images are also factors that affect computation time. When the sizes of the images are reduced by the resizing process, the time required for analyzing them decreases, and therefore, the performance of the HAR system is positively affected.

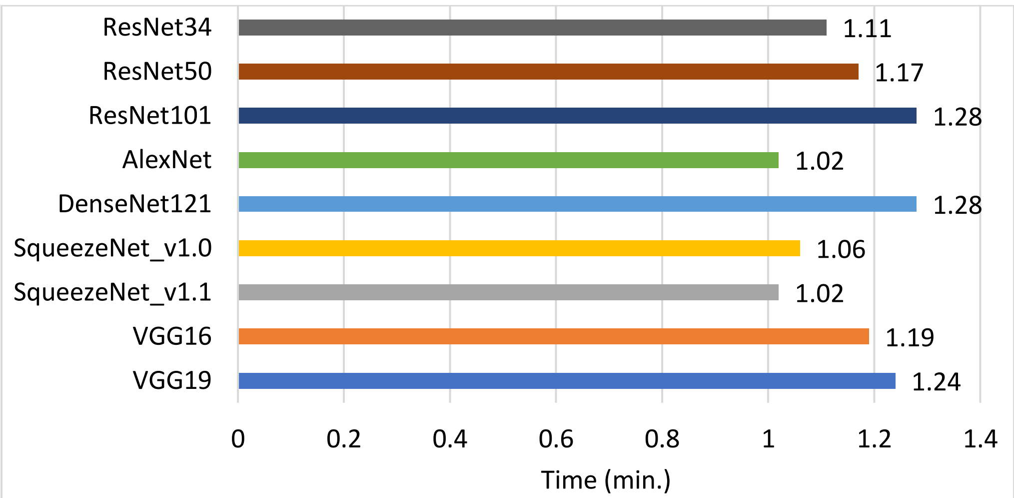

Figure 5.3 The execution time of HARSI method on different CNN architecture

CHAPTER SIX
DISCUSSION

6.1 Comparison with the Classical ML Methods

In order to show the superiority of our method, we compared it with the classical machine learning methods such as support vector machines (SVM), multi-layer perceptron (MLP), decision tree (DT), logistic regression (LR), naive Bayes (NB), k-nearest neighbors (KNN), AdaBoost, and random forest (RF). Table 6.1 presents the related works, which used the WISDM dataset (Kwapisz et al., 2011), along with the methods and the corresponding accuracy rates. It can be concluded from Table 6.1 that the proposed HARSI method outperformed the other methods (Bhat & Khan, 2021, Suwannarat & Kurdthongmee, 2021, Kolosnjaji & Eckert, 2015) with a 13.72% improvement on average. Employing HARSI achieved higher accuracy (98%) than the traditional machine learning models on the same dataset.

Table 6.1 Comparison of HARSI with the classical ML methods

Reference	**Year**	**Method**	**Accuracy (%)**
Bhat and Khan	2022	Support Vector Machines	87.40
		Random Forest	86.10
Suwannarat and Kurdthongmee	2021	Decision Tree	82.00
		Logistic Regression	68.00
		Multi-Layer Perceptron	80.00
		Neural Networks	94.00

Table 6.1 Continues

Vijayvargiya et al.	2021	Decision Tree	89.76
		Linear Discriminant Analysis	86.64
		Gradients Boosting	89.65
		K-Nearest Neighbors	92.54
		Bagging	92.48
		Random Forest	92.71
		Linear Kernel SVM	78.55
		RBF Kernel SVM	89.07
		Polynomial Kernel SVM	92.48
Semwal et al.	2021	Random Forest	79.38
		K-Nearest Neighbors	75.04
		Decision Tree	77.60
		Gradient Boosting	74.80
Kee et al.	2020	Random Forest	83.35
		Neural Networks	77.02
		Decision Tree (J48)	75.96
		Reduced-Error Pruning (REP) Tree	74.64
		K-Nearest Neighbors	72.08
		KStar	71.84
		Naive Bayes	63.89
		Support Vector Machines	55.45
Jalal et al.	2020	Random Forest	92.78
		Support Vector Machines	91.39

Table 6.1 Continues

Khare et al.	2020	Neural Networks	89.10
		Decision Tree	87.45
		Support Vector Machines	95.13
		Linear Support Vector Classifier	86.20
		Logistic Regression	81.10
		Random Forest	82.10
Arigbabu	2020	Support Vector Machines	82.00
Xu et al.	2020	Multi-Layer Perceptron	86.95
Lu et al.	2019	K-Nearest Neighbors	92.00
		Support Vector Machines	93.50
		Bagging	93.80
Ignatov	2018	Random Forest	82.66
		K-Nearest Neighbors	66.19
Xu et al.	2018	Support Vector Machines	82.27
Quispe et al.	2018	Naive Bayes	80.12
		Decision Tree	81.02
		K-Nearest Neighbors	77.58
		Support Vector Machines	80.93
Azmi and Sulaiman	2017	Naive Bayes Tree	87.70
		Multi-Layer Perceptron	77.52
		DT + LR + MLP	91.62
		NB Tree + MLP	96.35

Table 6.1 Continues

Walse et al.	2016	AdaBoost+J48	97.83
		AdaBoost + REP Tree	97.33
		AdaBoost + Random Tree	95.69
		AdaBoost + Random Forest	94.44
		AdaBoost + Hoeffding Tree	87.84
		AdaBoost + Decision Stump	57.31
Catal et al.	2015	Decision Tree (J48)	86.08
		Logistic Regression	77.52
		Multi-Layer Perceptron	88.81
		J48 + LR + MLP	91.62
Zainudin et al.	2015	Decision Tree (J48)	92.40
		Logistic Regression	84.30
		Multi-Layer Perceptron	91.70
		J48 + LR + MLP	93.00
Kolosnjaji and Eckert	2015	Neural Networks with Dropout	85.36
		Random Forest	83.46
Kwapisz et al.	2010	Logistic Regression	78.10
		Decision Tree (J48)	85.10
		Multi-Layer Perceptron	91.70
Our Approach		Human Activity Recognition on Signal Images (**HARSI**)	**98.00**

6.2 Comparison with the State-of-The-Art Methods

Table 6.2 shows the performance improvement of our method over the state-of-the-art methods (Mohsen et al., 2021, Bhat & Khan, 2021, Bozkurt, 2021, Alawneh et al., 2021, Ghate & C, 2021, Suwannarat & Kurdthongmee, 2021, Zhang et al., 2020, Abdallah et al., 2015) in the literature. The results were taken directly from the referenced studies since the researchers used the same dataset (Kwapisz et al., 2011) with our study. It can be concluded from the Table 6.2 that the proposed HARSI method outperformed the other methods with a 7.06% improvement on average.

Table 6.2 Comparison of HARSI with the state-of-the-art methods

Reference	**Year**	**Method**	**Accuracy (%)**
Bhat and Khan	2022	Convolutional Neural Networks - Transfer Learning	90.40
		Convolutional Neural Networks	88.20
Mohsen et al.	2021	Convolutional Neural Networks + Long Short-Term Memory	97.76
		Long Short-Term Memory	96.61
		Convolutional Neural Networks	94.51
Bozkurt	2021	Deep Neural Networks	93.00
Alawneh et al.	2021	Vanilla RNN + LSTM + GRU	97.13
Ghate and Hemalatha	2021	Convolutional Neural Networks + Random Forest	97.77
		Deep Neural Networks	74.00
		Deep Neural Networks + LSTM	81.00
		Deep Neural Networks + Gated Recurrent Unit (GRU)	80.00
		Convolutional Neural Networks	88.00
		Convolutional Neural Networks + LSTM	94.00
		Convolutional Neural Networks + GRU	82.00
Suwannarat and Kurdthongmee	2021	Deep Neural Networks	95.00

Table 6.2 Continues

Zhang et al.	2021	Residual Network	95.66
		Convolutional Neural Networks	92.19
Lin et al.	2021	Deep Convolutional Neural Networks (DCNN)	91.25
Zihao et al.	2021	1D Convolutional Neural Networks	91.12
		1D Convolutional Neural Networks + Fuzzy Neural Network	92.96
Garcia et al.	2021	Ensemble of Auto-Encoders (EAE)	82.00
		KNN + Very Fast Decision Tree + Naive Bayes (EkVN)	73.00
Lima et al.	2021	NOvelty discrete data stream for Human Activity Recognition (NOHAR)	93.00
Sena et al.	2021	Deep Convolutional Neural Networks Ensemble	89.01
Ramesh et al.	2021	Convolutional AutoEncoder (CAE)	95.60
Dhammi and Tewari	2021	Convolutional Neural Networks	95.00
		Long Short-Term Memory	97.50
Wenzheng	2020	Convolutional Neural Networks	93.25
Peppas et al.	2020	Deep Convolutional Neural Networks	94.18
		Region-based CNN	93.68
Mehmood et al.	2020	Convolutional Neural Networks - DenseNet	94.65
Aswal et al.	2020	Bidirectional Long Short-Term Memory	94.10
Jalal et al.	2020	Genetic algorithm-based classifier	95.37
Khare et al.	2020	Convolutional Neural Networks	83.98
		Long Short-Term Memory	95.45
Xia et al.	2020	LSTM- Convolutional Neural Networks	95.75

Table 6.2 Continues

Agarwal and Alam	2020	Lightweight Recurrent Neural Network - LSTM	95.78
Zhang et al.	2020	Multi-head Convolutional Attention	95.40
Huang et al.	2020	Two-Stage End-to-end CNN with data augmentation (TSE+CNN+Aug)	95.70
Xu et al.	2020	Gramian Angular Field + Multi-Dilated Kernel Residual Network	96.83
		Long Short-Term Memory	87.53
		1D Convolutional Neural Network	93.66
Mukherjee et al.	2020	EnsemConvNet (CNN-Net + Encoded-Net + CNN-LSTM)	97.20
Tang et al.	2020	Convolutional Neural Networks	97.51
Beirami and Shojaedini	2020	Convolutional Neural Networks	94.11
		Residual Network	95.72
		Residual Network of Residual Network (ROR)	96.73
Arigbabu	2020	Convolutional Neural Networks	81.70
		Recurrent Convolutional Network (RCN)	94.00
		Recurrent Convolutional Network + SVM	91.50
Zhang et al.	2019	U-Net	96.40
		Mask Region-based CNN (R-CNN)	86.20
		SegNet: A Deep Convolutional Encoder-Decoder Arch.	95.70
		Full Convolutional Network (FCN)	87.90
		Deep convolutional and LSTM (CovLSTM)	94.80
		Long Short-Term Memory	93.80
		Convolutional Neural Networks	94.10

Table 6.2 Continues

Pienaar and Malekian	2019	Long Short-Term Memory-Recurrent Neural Networks	93.81
Lu et al.	2019	Supervised Regularization-based Robust Subspace (SRRS)	93.50
		Robust Principal Component Analysis (RPCA)	85.70
		Latent Low-Rank Representation (LLRR)	91.90
		Joint Embedding Learning and Sparse Regression (JELSR)	73.40
		Principal Component Analysis (PCA)	92.30
		Linear Discriminant Analysis (LDA)	71.50
Manu et al.	2019	Long Short-Term Memory	97.00
Xu et al.	2018	Convolutional Neural Networks	91.97
Quispe et al.	2018	Multivariate Bag-Of-SFA-Symbols (MBOSS)	83.35
Varamin et al.	2018	Deep Auto-Encoder-Set Network	94.90
Chandini	2018	Long Short-Term Memory	97.00
Ignatov	2018	Convolutional Neural Networks	93.32
Dungkaew et al.	2017	Impersonal Smartphone-based Activity Recognition (ISAR)	75.21
Chen et al.	2016	Long Short-Term Memory	92.10
Abdallah et al.	2015	Stream learning for mobile Activity Recognition (STAR)	71.20
Our Approach		Human Activity Recognition on Signal Images (**HARSI**)	**98.00**

CHAPTER SEVEN
CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

Standard HAR has been defined as a time-series data classification problem and uses feature extraction data, which cannot be readable and interpretable by humans. Transparent and explainable HAR systems are required that can generate human-understandable information. For this purpose, an approach, called Human Activity Recognition on Signal Images (HARSI), is proposed in this study. The proposed approach creates image representations of the time series sensor data to improve both explainability and recognition accuracy. It takes advantage of the strengths of CNNs in handling image data. In the experimental studies, we demonstrated the effectiveness of the proposed HARSI approach on a real-world dataset.

The main debates in the field of HAR and our solutions can be summarized as:

- In HAR, the ideal input data format is still a subject of much debate and there are various ongoing works for improving the accuracy of the models. Traditional HAR has been defined as a time-series data classification problem and requires feature extraction. In contrast, we transfer time-series data into signal images that reflect the properties of human activities. It avoids the need to perform an explicit feature generation and selection stage. We improved accuracy by working on signal-image data, instead of numerical time-series data.
- Many applications in HAR have used classical machine learning methods such as DT, SVM, MLP, NB, LR, KNN, and RF. However, the performance of these methods is still highly debated. In this study, we take advantage of the strengths of deep learning approaches.
- Another debate is how to design CNN architecture to be able to obtain good performance. For example, the number of layers and parameter settings are still subjects of much debate. In this study, we compared nine different CNN architectures to determine the best suitable one.
- In the activity recognition community, there is an open debate on providing explainability in the HAR systems. The main problem is how to increase the transparency and interpretability of the models. In this study, to increase human-level

explainability, we visualize the data with charts since generating signal images makes data understandable for humans.

- Another ongoing debate is which activities can be predicted more precisely. This study showed that the best accuracies were achieved in the sitting, standing, and jogging activities due to their diverse natures.

The proposed HAR model can be connected to many different fields of study such as health monitoring, fitness tracking, home and work automation, and self-managing system. With the rapid technological developments in smartphones, the model can enable new opportunities for developing informative systems on a large scale to perceive and act on what users (i.e., your children, elderly mother, or sick family member) are doing. Recognizing human activities is important for the treatment of patients which can provide useful feedback to clinicians since the activity is associated with health. For example, it can be used to monitor patients in rehabilitation since the functional status of a person is an important parameter in this area. Besides, it could be used to offer activity-aware services to the smartphone users like movement recommendations. Several lifestyle diseases and movement disorders are associated with inactivity and therefore, our model can be used to give information to prevent them from diseases. The users can participate in the tracking of their activities for the sake of health, fitness, or other purposes due to its strength in providing personalized support.

7.2 Main Findings

The main findings of the study can be concluded as follows:

- The proposed approach improves human-level explainability for smart sensor data by using signal images in the field of HAR.
- The proposed HARSI approach improves the recognition accuracy in the HAR problems by converting time-series data to image data.
- The experimental results showed that HARSI successfully (98%) recognized six daily human activities, including walking, jogging, standing, sitting, downstairs, and upstairs.
- According to the experimental results, it can be concluded that the best suitable and consistent CNN model for the WISDM dataset is VGG19. It achieved the best

results on all the metrics (accuracy, precision, recall, and f-measure). Therefore, this model can be successfully used to identify human activities.

- The prediction accuracy changes according to human activities. Among the activities, sitting, standing, and jogging were correctly predicted by the proposed method. On the other hand, the model had a little difficulty in classifying downstairs and upstairs activities with an accuracy of 95% for the WISDM dataset.
- The accuracy of image classification changes slightly depending on the quality (resolution) of the signal images.
- A significant improvement (13.72% on average) was achieved by the proposed HARSI model compared to the classical machine learning methods such as KNN, DT, SVM, NB, LR, MLP, AdaBoost, and RF.
- Our approach achieved higher classification accuracy than the state-of-the-art approaches. It outperformed them by 7.06% on average on the same dataset.
- The proposed HARSI approach has the potential to expand the application of machine learning in many different sectors, thanks to its advantages.

7.3 Future Work

Future work can be focused on using the proposed approach for constructing new models, including more types of activities. Moreover, the impact of smartphone location on the performance of the proposed approach can be investigated in a future study.

One limitation of this study is related to sensors such as signal delays, noises, damages, battery capacity, and shelf-life. However, this limitation is also valid for all other wearable sensor-based HAR applications. It can be overcome in the future with developments in sensor technology. Another limitation is that it focuses on single-person activity detection. In the future, we plan to adapt it for recognizing group activities such as handshaking and hugging.

REFERENCES

Abdallah, Z. S., Gaber, M. M., Srinivasan, B., & Krishnaswamy, S. (2015). Adaptive Mobile Activity Recognition System with evolving data streams. *Neurocomputing*, *150*, 304–317. https://doi.org/10.1016/j.neucom.2014.09.074

Agarwal, P., & Alam, M. (2020). A lightweight deep learning model for human activity recognition on edge devices. *Procedia Computer Science*, *167*, 2364–2373. https://doi.org/10.1016/j.procs.2020.03.289

Al-Wesabi, F. N., Albraikan, A. A., Hilal, A. M., Al-Shargabi, A. A., Alhazbi, S., Al Duhayyim, M., Rizwanullah, M., & Hamza, M. A. (2021). Design of Optimal Deep Learning based human activity recognition on sensor enabled internet of things environment. *IEEE Access*, *9*, 143988–143996. https://doi.org/10.1109/access.2021.3112973

Alawneh, L., Alsarhan, T., Al-Zinati, M., Al-Ayyoub, M., Jararweh, Y., & Lu, H. (2021). Enhancing human activity recognition using Deep Learning and time series augmented data. *Journal of Ambient Intelligence and Humanized Computing*, *12*(12), 10565–10580. https://doi.org/10.1007/s12652-020-02865-4

Alhersh, T., Stuckenschmidt, H., Rehman, A. U., & Belhaouari, S. B. (2021). Learning human activity from visual data using Deep Learning. *IEEE Access*, *9*, 106245–106253. https://doi.org/10.1109/access.2021.3099567

Arigbabu, O. A. (2020). Entropy Decision Fusion for Smartphone Sensor based Human Activity Recognition. doi:10.48550/ARXIV.2006.00367

Aswal, V., Sreeram, V., Kuchik, A., Ahuja, S., & Patel, H. (2020). Real-time human activity generation using bidirectional long short-term memory networks. *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*. https://doi.org/10.1109/iciccs48265.2020.9121096

Baysari, K., Birant, D., & Cengiz, A. (2022). Human Activity Recognition using machine learning methods. *2022 7th International Battal Gazi Natural Scientific Congress,* 643–653.

Beirami MJ, & Shojaedini SV. (2020).Residual network of residual network: a new deep learning modality to improve human activity recognition by using smart sensors exposed to unwanted shocks. *J Health Man Inf 20207(4):*228-39

Bhat, O., & Khan, D. A. (2021). Evaluation of deep learning model for human activity recognition. *Evolving Systems*, *13*(1), 159–168. https://doi.org/10.1007/s12530-021-09373-6

Bijalwan, V., Semwal, V. B., & Gupta, V. (2021). Wearable sensor-based pattern mining for human activity recognition: Deep Learning Approach. *Industrial Robot: the International Journal of Robotics Research and Application,* *49*(1), 21–33. https://doi.org/10.1108/ir-09-2020-0187

Bozkurt, F. (2021). A comparative study on classifying human activities using classical machine and deep learning methods. *Arabian Journal for Science and Engineering*, *47*(2), 1507–1521. https://doi.org/10.1007/s13369-021-06008-5

Buffelli, D., & Vandin, F. (2021). Attention-based deep learning framework for human activity recognition with user adaptation. *IEEE Sensors Journal*, *21*(12), 13474–13483. https://doi.org/10.1109/jsen.2021.3067690

Catal, C., Tufekci, S., Pirmit, E., & Kocabag, G. (2015). On the use of ensemble of classifiers for accelerometer-based activity recognition. *Applied Soft Computing*, *37*, 1018–1022. https://doi.org/10.1016/j.asoc.2015.01.025

Chandini U. (2018). A Machine learning based activity recognition for ambient assisted living, *International Journal on Future Revolution in Computer Science & Communication Engineering. 2018;4*(3*)*:323-26.

Chen, L., Liu, X., Peng, L., & Wu, M. (2020). Deep Learning based multimodal complex human activity recognition using wearable devices. *Applied Intelligence*, *51*(6), 4029–4042. https://doi.org/10.1007/s10489-020-02005-7

Chen, Y., Zhong, K., Zhang, J., Sun, Q., & Zhao, X. (2016). LSTM networks for mobile human activity recognition. *Proceedings of the 2016 International Conference on Artificial Intelligence: Technologies and Applications*. https://doi.org/10.2991/icaita-16.2016.13

Cengiz, A. B., Birant, K. U., Cengiz, M., Birant, D., & Baysari, K. (2022). Improving the performance and explainability of indoor human activity recognition in the internet of things environment. *Symmetry*, *14*(10), 2022. https://doi.org/10.3390/sym14102022

Dhammi, L., & Tewari, P. (2021). Classification of human activities using data captured through a smartphone using Deep Learning Techniques. *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*. https://doi.org/10.1109/icspc51351.2021.9451772

Dungkaew, T., Suksawatchon, J., & Suksawatchon, U. (2017). Impersonal smartphone-based activity recognition using the accelerometer sensory data.

2017 2nd International Conference on Information Technology (INCIT). https://doi.org/10.1109/incit.2017.8257856

Elsts, A., & McConville, R. (2021). Are microcontrollers ready for deep learning-based human activity recognition? *Electronics*, *10*(21), 2640. https://doi.org/10.3390/electronics10212640

Ferrari, A., Micucci, D., Mobilio, M., & Napoletano, P. (2022). Deep learning and model personalization in sensor-based human activity recognition. *Journal of Reliable Intelligent Environments*, *2022:1-13*. https://doi.org/10.1007/s40860-021-00167-w

Garcia, K. D., de Sá, C. R., Poel, M., Carvalho, T., Mendes-Moreira, J., Cardoso, J. M. P., de Carvalho, A. C. P. L. F., & Kok, J. N. (2021). An ensemble of autonomous auto-encoders for human activity recognition. *Neurocomputing*, *439*, 271–280. https://doi.org/10.1016/j.neucom.2020.01.125

Ghate, V., & C, S. H. (2021). Hybrid deep learning approaches for smartphone sensor-based human activity recognition. *Multimedia Tools and Applications*, *80*(28-29), 35585–35604. https://doi.org/10.1007/s11042-020-10478-4

Hassan, M. M., Ullah, S., Hossain, M. S., & Alelaiwi, A. (2020). An end-to-end deep learning model for human activity recognition from highly sparse body sensor data in internet of medical things environment. *The Journal of Supercomputing*, *77*(3), 2237–2250. https://doi.org/10.1007/s11227-020-03361-4

Hoai Thu, N. T., & Han, D. S. (2021). Hihar: A hierarchical hybrid deep learning architecture for wearable sensor-based human activity recognition. *IEEE Access*, *9*, 145271–145281. https://doi.org/10.1109/access.2021.3122298

Huang, J., Lin, S., Wang, N., Dai, G., Xie, Y., & Zhou, J. (2020). Tse-CNN: A two-stage end-to-end CNN for human activity recognition. *IEEE Journal of Biomedical and Health Informatics*, *24*(1), 292–299. https://doi.org/10.1109/jbhi.2019.2909688

Hwang, Y. M., Park, S., Lee, H. O., Ko, S.-K., & Lee, B.-T. (2021). Deep learning for human activity recognition based on causality feature extraction. *IEEE Access*, *9*, 112257–112275. https://doi.org/10.1109/access.2021.3103211

Ignatov, A. (2018). Real-time human activity recognition from accelerometer data using convolutional neural networks. *Applied Soft Computing*, *62*, 915–922. https://doi.org/10.1016/j.asoc.2017.09.027

Irfan, S., Anjum, N., Masood, N., Khattak, A. S., & Ramzan, N. (2021). A novel hybrid deep learning model for human activity recognition based on transitional activities. *Sensors, 21*(24), 8227. https://doi.org/10.3390/s21248227

Jalal, A., Quaid, M. A., Tahir, S. B., & Kim, K. (2020). A study of accelerometer and gyroscope measurements in physical life-log activities detection systems. *Sensors*, *20*(22), 6670. https://doi.org/10.3390/s20226670

Kanjilal, R., & Uysal, I. (2021). The future of human activity recognition: Deep Learning or feature engineering? *Neural Processing Letters*, *53*(1), 561–579. https://doi.org/10.1007/s11063-020-10400-x

Kee, Y. J., Shah Zainudin, M. N., Idris, M. I., Ramlee, R. H., & Kamarudin, M. R. (2020). Activity recognition on subject independent using machine learning. *Cybernetics and Information Technologies, 20*(3), 64–74. https://doi.org/10.2478/cait-2020-0028

Khan, A. R., Saba, T., Khan, M. Z., Fati, S. M., & Khan, M. U. (2022). Classification of human's activities from gesture recognition in live videos using Deep Learning. *Concurrency and Computation: Practice and Experience*, *34*(10). https://doi.org/10.1002/cpe.6825

Khan, I. U., Afzal, S., & Lee, J. W. (2022). Human activity recognition via hybrid deep learning based model. *Sensors*, *22*(1), 323. https://doi.org/10.3390/s22010323

Khare, S., Sarkar, S., & Totaro, M. (2020). Comparison of sensor-based datasets for human activity recognition in wearable IOT. *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*. https://doi.org/10.1109/wf-iot48130.2020.9221408

Kolosnjaji, B., & Eckert, C. (2015). Neural network-based user-independent physical activity recognition for mobile devices. *Intelligent Data Engineering and Automated Learning – IDEAL 2015*, 378–386. https://doi.org/10.1007/978-3-319-24834-9_44

Kothamachu Ramesh, A., Gajjala, K. S., Nakano, K., & Chakraborty, B. (2021). Person authentication by gait data from smartphone sensors using convolutional Autoencoder. *Intelligence Science III*, 149–158. https://doi.org/10.1007/978-3-030-74826-5_13

Kwapisz, J. R., Weiss, G. M., & Moore, S. A. (2011). Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations Newsletter*, *12*(2), 74–82. https://doi.org/10.1145/1964897.1964918

Lima, W. S., Bragança, H. L. S., & Souto, E. J. P. (2021). Nohar - novelty discrete data stream for human activity recognition based on smartphones with inertial sensors. *Expert Systems with Applications*, *166*, 114093. https://doi.org/10.1016/j.eswa.2020.114093

Lin SB, Wang K, Wang Y, Zhou DX. (2021). *Universal consistency of deep convolutional neural networks. 2021*, arXiv preprint *arXiv:2106.12498*.

Lu, W., Fan, F., Chu, J., Jing, P., & Yuting, S. (2019). Wearable computing for internet of things: A discriminant approach for human activity recognition. *IEEE Internet of Things Journal*, *6*(2), 2749–2759. https://doi.org/10.1109/jiot.2018.2873594

Madokoro, H., Nix, S., Woo, H., & Sato, K. (2021). A mini-survey and feasibility study of deep-learning-based human activity recognition from slight feature signals obtained using privacy-aware environmental sensors. *Applied Sciences*, *11*(24), 11807. https://doi.org/10.3390/app112411807

Maitre, J., Bouchard, K., & Gaboury, S. (2021). Alternative deep learning architectures for feature-level fusion in human activity recognition. *Mobile Networks and Applications*, *26*(5), 2076–2086. https://doi.org/10.1007/s11036-021-01741-5

Manjarres, J., Lan, G., Gorlatova, M., Hassan, M., & Pardo, M. (2022). Deep learning for detecting human activities from piezoelectric-based kinetic energy signals. *IEEE Internet of Things Journal*, *9*(10), 7545–7558. https://doi.org/10.1109/jiot.2021.3093245

Manu RD, Kumar S, Snehashish S, Rekha KS. (2019). Smart home automation using IoT and deep learning. *Int Res J Eng Techn International Journal of Scientific & Engineering Research. 11. 697-701. 10.14299/ijser.2020.06.03. 2019*;6(4):1-4.

Mehmood, K., Imran, H. A., & Latif, U. (2020). Hardensenet: A 1d DenseNet inspired convolutional neural network for human activity recognition with inertial sensors. *2020 IEEE 23rd International Multitopic Conference (INMIC)*. https://doi.org/10.1109/inmic50486.2020.9318067

Mekruksavanich, S., & Jitpattanakul, A. (2021). Biometric user identification based on human activity recognition using wearable sensors: An experiment using Deep Learning Models. *Electronics*, *10*(3), 308. https://doi.org/10.3390/electronics10030308

Mihoub, A. (2021). A deep learning-based framework for human activity recognition in Smart Homes. *Mobile Information Systems*, *2021*, 1–11. https://doi.org/10.1155/2021/6961343

Mohd Azmi, M. S., & Sulaiman, M. N. (2017). Accelerator-based human activity recognition using voting technique with NBTree and MLP classifiers. *International Journal on Advanced Science, Engineering and Information Technology*, *7*(1), 146. https://doi.org/10.18517/ijaseit.7.1.1790

Mohsen, S., Elkaseer, A., & Scholz, S. G. (2021). Industry 4.0-oriented deep learning models for human activity recognition. *IEEE Access*, *9*, 150508–150521. https://doi.org/10.1109/access.2021.3125733

Montero Quispe, K., Sousa Lima, W., Macêdo Batista, D., & Souto, E. (2018). Mboss: A symbolic representation of human activity recognition using mobile sensors. *Sensors*, *18*(12), 4354. https://doi.org/10.3390/s18124354

Mukherjee, D., Mondal, R., Singh, P. K., Sarkar, R., & Bhattacharjee, D. (2020). EnsemConvNet: A deep learning approach for human activity recognition using smartphone sensors for healthcare applications. *Multimedia Tools and*

Applications, *79*(41-42), 31663–31690. https://doi.org/10.1007/s11042-020-09537-7

Pei, L., Xia, S., Chu, L., Xiao, F., Wu, Q., Yu, W., & Qiu, R. (2021). Mars: Mixed virtual and real wearable sensors for human activity recognition with Multidomain Deep Learning Model. *IEEE Internet of Things Journal*, *8*(11), 9383–9396. https://doi.org/10.1109/jiot.2021.3055859

Peppas, K., Tsolakis, A. C., Krinidis, S., & Tzovaras, D. (2020). Real-time physical activity recognition on smart mobile devices using Convolutional Neural Networks. *Applied Sciences*, *10*(23), 8482. https://doi.org/10.3390/app10238482

Pienaar, S. W., & Malekian, R. (2019). Human activity recognition using LSTM-RNN deep neural network architecture. *2019 IEEE 2nd Wireless Africa Conference (WAC)*. https://doi.org/10.1109/africa.2019.8843403

Ronald, M., Poulose, A., & Han, D. S. (2021). ISPLInception: An inception-resnet deep learning architecture for human activity recognition. *IEEE Access*, *9*, 68985–69001. https://doi.org/10.1109/access.2021.3078184

Semwal, V. B., Lalwani, P., Mishra, M. K., Bijalwan, V., & Chadha, J. S. (2021). An optimized feature selection using bio-geography optimization technique for human walking activities recognition. *Computing*, *103*(12), 2893–2914. https://doi.org/10.1007/s00607-021-01008-7

Sena, J., Barreto, J., Caetano, C., Cramer, G., & Schwartz, W. R. (2021). Human activity recognition based on smartphone and wearable sensors using multiscale DCNN ensemble. *Neurocomputing*, *444*, 226–243. https://doi.org/10.1016/j.neucom.2020.04.151

Shalaby, E., ElShennawy, N., & Sarhan, A. (2022). Utilizing deep learning models in CSI-based human activity recognition. *Neural Computing and Applications*, *34*(8), 5993–6010. https://doi.org/10.1007/s00521-021-06787-w

Stuart, M., & Manic, M. (2021). Deep learning shared bandpass filters for resource-constrained human activity recognition. *IEEE Access*, *9*, 39089–39097. https://doi.org/10.1109/access.2021.3064031

Suwannarat, K., & Kurdthongmee, W. (2021). Optimization of deep neural network-based human activity recognition for a wearable device. *Heliyon*, *7*(8). https://doi.org/10.1016/j.heliyon.2021.e07797

Tang Y, Teng Q, Zhang L, Min F, He J. Efficient convolutional neural networks with smaller filters for human activity recognition using wearable sensors. 2020, arXiv preprint arXiv:2005.03948v1.

Tasnim, N., Islam, M. K., & Baek, J.-H. (2021). Deep learning based human activity recognition using spatio-temporal image formation of skeleton joints. *Applied Sciences*, *11*(6), 2675. https://doi.org/10.3390/app11062675

Thakur, D., & Biswas, S. (2021). Feature fusion using Deep Learning for Smartphone based human activity recognition. *International Journal of Information Technology*, *13*(4), 1615–1624. https://doi.org/10.1007/s41870-021-00719-6

Varamin, A. A., Abbasnejad, E., Shi, Q., Ranasinghe, D. C., & Rezatofighi, H. (2018). Deep auto-set. *Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. https://doi.org/10.1145/3286978.3287024

Vijayvargiya, A., Kumari, N., Gupta, P., & Kumar, R. (2021). Implementation of machine learning algorithms for human activity recognition. *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*. https://doi.org/10.1109/icspc51351.2021.9451802

Walse KH, Dharaskar RV, Thakare VM. (2016). A study of human activity recognition using adaboost classifiers on wisdm dataset. *Inst Integr Omics Appl Biotechnol, The IIOAB JournalJ. 2016*;7(2):68-76.

Wan, S., Qi, L., Xu, X., Tong, C., & Gu, Z. (2019). Deep learning models for real-time human activity recognition with smartphones. *Mobile Networks and Applications*, *25*(2), 743–755. https://doi.org/10.1007/s11036-019-01445-x

Wenzheng, Z. (2020). Human activity recognition based on acceleration sensor and Neural Network. *2020 8th International Conference on Orange Technology (ICOT)*. https://doi.org/10.1109/icot51877.2020.9468785

Wikimedia Foundation. (2022, September 2). *Deep learning*. Wikipedia. Retrieved September 7, 2022, from https://en.wikipedia.org/wiki/Deep_learning

Xia, K., Huang, J., & Wang, H. (2020). LSTM-CNN Architecture for Human Activity Recognition. *IEEE Access*, *8*, 56855–56866. https://doi.org/10.1109/access.2020.2982225

Xu, H., Li, J., Yuan, H., Liu, Q., Fan, S., Li, T., & Sun, X. (2020). Human activity recognition based on Gramian angular field and deep convolutional neural network. *IEEE Access*, *8*, 199393–199405. https://doi.org/10.1109/access.2020.3032699

Xu, W., Pang, Y., Yang, Y., & Liu, Y. (2018). Human activity recognition based on Convolutional Neural Network. *2018 24th International Conference on Pattern Recognition (ICPR)*. https://doi.org/10.1109/icpr.2018.8545435

Yen, C.-T., Liao, J.-X., & Huang, Y.-K. (2021). Feature fusion of a deep-learning algorithm into wearable sensor devices for human activity recognition. *Sensors*, *21*(24), 8294. https://doi.org/10.3390/s21248294

Zainudin, M. N. S., Sulaiman, M. N., Mustapha, N., & Perumal, T. (2015). Activity recognition based on accelerometer sensor using Combinational classifiers. *2015 IEEE Conference on Open Systems (ICOS)*. https://doi.org/10.1109/icos.2015.7377280

Zhang, H., Xiao, Z., Wang, J., Li, F., & Szczerbicki, E. (2020). A novel IOT-perceptive human activity recognition (har) approach using multihead convolutional attention. *IEEE Internet of Things Journal*, *7*(2), 1072–1080. https://doi.org/10.1109/jiot.2019.2949715

Zhang, J., Qiao, S., Lin, Z., & Zhou, Y. (2021). Human activity recognition based on residual network. *IOP Conference Series: Earth and Environmental Science*, *693*(1), 012041. https://doi.org/10.1088/1755-1315/693/1/012041

Zhang, Y., Zhang, Z., Zhang, Y., Bao, J., Zhang, Y., & Deng, H. (2019). Human activity recognition based on motion sensor using U-Net. *IEEE Access*, *7*, 75213–75226. https://doi.org/10.1109/access.2019.2920969

Zihao, Z., Geng, J., & Jiang, W. (2021). A time series classification method based on 1DCNN-FNN. *2021 33rd Chinese Control and Decision Conference (CCDC)*. https://doi.org/10.1109/ccdc52312.2021.9602164