Activity Detection Systems Using Infrared Array Sensors With Deep Learning

August 2022

Krishnan Arumugasamy Muthukumar

Activity Detection Systems Using Infrared Array Sensors With Deep Learning

by

Krishnan Arumugasamy Muthukumar

Submitted to the Graduate School of Science and Technology
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Ph.D) in Engineering

at the

KEIO UNIVERSITY

August 2022

Author .....
Graduate School of Science and Technology
August 2022

Certified by .....
Tomoaki Ohtsuki
Professor, Keio University, Supervisor

Certified by .....
Masaaki Ikehara
Professor, Keio University, Co-supervisor

Certified by .....
Hideo Saito
Professor, Keio University, Co-supervisor

Certified by .....
Guan Gui
Professor, NJUPT, Co-supervisor

Acknowledgments

Undertaking this PhD has been a truly life-changing experience. It would not have been possible to do without the support, guidance and encouragement of many people.

First and foremost, I would like to express my deepest and sincere gratitude to my supervisor Prof. Tomoaki Ohtsuki for the continuous support and encouragement that he gave me. Without his patience, motivation and immense knowledge, this PhD would not have been achievable. His advice was crucial to undertake new research challenges and keep persevering even in hard times when results were hard to obtain. I am deeply grateful for all the empowerment I received under his supervision that let me define my own pace.

The committee members Prof. Masaaki Ikehara, Prof. Hideo Saito and Prof. Guan Gui, deserve a special mention, for their precious time and advice that helped improve the quality of this dissertation.

Assistance provided by The Ministry of Education, Culture, Sports, Science and Technology (Monbukagakusho: MEXT) and Keio Leading Edge Laboratory (KLL), by offering grants and scholarship to support research during my PhD study has been of a great help and deserve a special thank you.

Many thanks to all the members of Ohtsuki Laboratory, especially Dr. Mondher Bouazizi who have always been there for me, have never let me work alone and have always known how to keep me enthusiastic and looking forward to facing new challenges. Their unlimited energy is a reference for my future endeavors.

Last but not least, I want to say a heartfelt thank to my family and friends for their love and support throughout my life. Thank you for giving me strength and support for undertaking this Ph.D.

Contents

List of Figures

List of Tables

Abstract

In assistive care technologies, activity detection is one of the vital tasks to assist people by preventing or at least detecting any accident that might occur. Activity detection has conventionally relied on two leading families of devices: wearable and non-wearable ones. As their name suggests, wearable devices are devices that require the person being monitored to wear them or at least carry them with him/her anywhere (s)he goes, such as smartphones, smartwatches, accelerometers, kinetic sensors, etc. It is a burden to the person to carry the device. Non-wearable devices, on the other hand, do not have such limiting constraints. A device (typically a sensor) is placed in a specific location in the area under monitoring, with no need for the monitored person to worry about its functioning. In recent years, many non-contact activity detection techniques have been proposed using Wireless Fidelity (Wi-Fi), Light Detection and Ranging (LiDAR), radar etc. These approaches have limitations like coverage issues, and deployment issues related to computational resources.

The recent introduction of the wide-angle low-resolution infrared (IR) array sensor helped develop device-free monitoring systems to solve most of the issues. Many IR-based activity detection systems have been proposed in recent years. The limitations of the existing works include but are not limited to the difficulty to detect the activity, the non-robustness to the environment, and the computational resource constraints in deployment.

To address the aforementioned issues, this thesis proposes activity detection systems using a hybrid deep learning model, which could classify blurriness and noisy images produced by the two wide-angle IR array sensors. One is placed on the wall, and another one is placed on the ceiling. Activity detection technique involves two stages. First, we classify the individual frames collected by the wall sensor and the ceiling sensor separately using a Convolution Neural Network (CNN). In the second stage, the output of the CNN is passed through a Long Short-Term Memory (LSTM) with a window size equal to 5 frames to classify the sequence of activities. Afterwards, we combine the ceiling data and wall data and classify each pair of frames using hy-

brid deep learning model. Furthermore, we propose an activity detection systems using one IR array sensor on the ceiling allowing for performance comparable to that when using dual sensors. By applying advance deep learning based computer vision techniques, we remove the noise and blurriness in data, which help to improve the IR image quality. The IR images/image sequences are then classified using a hybrid DL model that combines a CNN and an LSTM. By incorporating a wider variety of samples, we use data augmentation to improve the training of neural networks and make the model robust to the environment. A Conditional Generative Adversarial Network (CGAN) performs the data augmentation process. By enhancing the images with Super-Resolution (SR), removing the noise, and augmenting the training data with more samples, the classification accuracy of activity detection can be improved. We used quantization to optimize the neural network so that it could run on low-powered devices.

The contribution of the thesis as follows:

- We propose a lightweight Deep Learning model for activity classification that is robust to environmental changes. Being lightweight, such a model can run on devices with very low computation capabilities, making it a base for a cheap solution for activity detection.

- The blurriness and noise present in the IR captured frames, due to the sensor characteristics the imprecision in the sensor lead to a noticeable drop in performance in conventional methods. Our proposed neural network architecture manages to address this issue by exploiting the temporal changes in the frames to identify the activities accurately.

- We identify the activity using a time window of less than 1 second. Despite the smaller time window, we have remarkably enhanced the classification accuracy in comparison to conventional works, which require a larger time window.

- Low Resolution (LR) sensors are always preferred over High Resolution (HR) ones if they provide similar performance. It preserve the privacy of the person

and have much lower cost. We demonstrate that it is possible, by using deep learning techniques such as Super-resolution, denoising, and CGAN, to achieve classification performance on the LR data that is nearly identical to that of the classification of the HR data, namely 24×32.

Chapter 1

Introduction

Population ageing is a societal issue facing many countries nowadays that affects not only social life but also the economy. As a matter of fact, advancements in healthcare and medicine have continuously increased the average life expectancy over the last few decades. Today, the total world population stands at 7.9 billion [1] with 703 million people above the age of 65. Asia and Europe account for most of the elderly population in the world. Japan, for instance, is at the very top, with 28% [2] of its population above the age of 65. This high ratio of elderly people and increase in life expectancy combined with the fact that most of these people are living alone have made it necessary to develop more sophisticated techniques and technologies to monitor them. In this regard, artificial intelligence (AI) [3] plays an important role in healthcare, particularly in assistive care technologies [4–6] for old people owing to the spurt in the Internet of Things (IoT)-based technological applications.

Activity detection (AD) is one of the vital positions in assistive care technologies for assisting individuals by preventing or at least detecting any accidents that may occur. In general, AD has relied on two major families of devices: wearable and non-wearable. Wearable devices, as the name implies, are devices that require the person being monitored to wear them or take them with him/her everywhere (s)he goes. Smartphone [7], smartwatch [8], accelerometer [9], and kinetic sensors [10] all fall within this category. Carrying these devices, users must be extremely careful and ensure that they are always taking the correct measurements (e.g., the smartwatch

is well placed, etc.). Elderly persons may not prefer such a burden because it causes them discomfort. Non-wearable devices, on the other hand, are not constrained by such limits. A device (usually a sensor) is placed in a precise location in the monitored area, with no requirement for the monitored person to be concerned about its operation. Several approaches were proposed in the literature to do so. They used array antennas [11], doppler radars [12], Light Detection and Ranging (LiDAR) [13], Wi-Fi [14], and infrared (IR) array sensors [15]. An example of the devices are shown in Fig. 1-1

Figure 1-1: Some examples of wearable and non-wearable devices.

Wearable sensor technology has the ability to sense, collect, and upload physiological data in a continuous manner, providing opportunities to improve quality of life in a way that is not easily attainable with smartphones alone. Additionally, wearable sensors can make it much easier and more natural for users to complete numerous other helpful micro tasks, such as checking incoming text messages and viewing urgent information, than is possible with a smartphone, which is frequently carried in pockets or bags. A variety of value-added services are also offered by wearables, including indoor localization and navigation [16], [17], financial payments [18], [19], monitoring

of physical and mental health [20], [21], sport analytics [22], and medical insurance analytics [23]. Recent market reports [24], predict a 44.4 percent increase in wearable device shipments for 2016, compared to the 80 million devices shipped in 2015, with annual shipments reaching 200 million by 2019. By 2022, the market for wearable technology is anticipated to reach a value of $57,653 million, nearly threefold that of 2016 ($19,633 million).

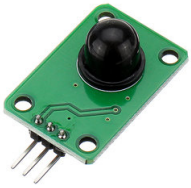
As they offer a combination of human activity recognition and vital sign detection, wearable sensor technologies are an efficient solution for smart healthcare applications [25–28]. Wearable sensors are capable of capturing small fractions of the body, such as finger movement [29]. Wearable sensing technologies can also detect physiological signals like heart rate and speech patterns [30]. However, the disadvantages of this technology are that the battery life of wearable sensors is notoriously short. Because they are "wearable," elderly people may forget to wear them or feel uncomfortable wearing them [31].

Non-wearable technology is independent of the user. They don't have to wear the device. There are many non- wearable based technologies available, including acoustic-based [32, 33], ambient, and vision-based sensors [34]. The acoustic-based [35] approaches is vulnerable to ambient noise and surrounding sound interference, and the sensing range is also limited due to the fast attenuation of acoustic signals. The ambient-based approaches are vulnerable to power constraint signal weakness, and high temperature caused data affected. The vision-based approaches relying on camera [36] or visible light sensors can only work well in environments under certain light conditions, which could be easily interfered with by low illumination conditions, smoke, or opaque obstructions. The better solution for those mentioned above issued is infrared imaging it is also one of the vision-based approaches. The thermal camera or sensor [37] is essentially a passive sensor that detects infrared variations emitted by people in a room. Initially, they were typically used for nighttime surveillance for military purposes [38, 39]. But since they have gotten much cheaper in recent years, they are now affordable and popular for everyday uses as well. Thermal cameras can detect the human body and motion in a scene regardless of variations in illumination,

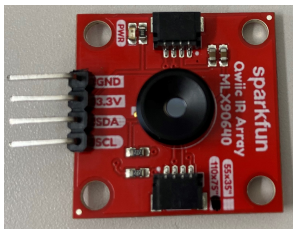
color of human surfaces, and backgrounds for human activity detection [40]. On the other hand, most normal cameras rely heavily on these parameters to create depth images of their surroundings. As a result, thermal cameras or sensors can be useful for monitoring human activities, particularly in low-light environments [41] where other conventional cameras would typically fail to provide enough data to accomplish the purpose of the various applications.

1.1 AD Based on IR Array Sensor

IR array sensors have attracted attention in healthcare technologies [42]. The IR array sensor measures the heat generated from the human body and projects it on a low-resolution (LR) matrix which could then be visualized as an image. It has several advantages: non-invasive from a privacy perspective, ease of positioning/set-up, better coverage resulting in a wider area of detection, etc. Moreover, its low cost makes it affordable to implement. These advantages make IR array sensors economical for use in a variety of industries such as aerospace [43], healthcare [44], automotive [45], etc. In the research related to the human AD, two types of IR sensors are used: pyroelectric infrared sensor (PIS) [46] and thermophile IR array sensor [47] [48] are shown in Fig. 1-2. The PIS is only capable of detecting motion-type of human activities (i.e., activity where the person or part of his body is moving). It is not capable of detecting static human activities (e.g., when he is sitting still, standing, or laying, etc.) [49]. The thermopile IR array sensor quantifies the temperature distribution within the field of view. It is capable of detecting static and dynamic human activities, providing us with an understanding of the surrounding environment and pertinent data. As a result, several studies have used such sensors for human AD, position detection, counting the number of people [50–54] in a room, etc.



Pyroelectric infrared sensor



Thermophile infrared sensor

Figure 1-2: Two types of IR sensor.

The IR sensor-based AD systems have relied mostly on two categories of classifiers: the conventional machine learning classifiers, and the more advanced Deep Learning (DL) ones. Machine learning models such as Support Vector Machine (SVM) [15], k-Nearest Neighbors (k-NN) [55], random forest [56], and others were used in conventional machine learning approaches. These conventional approaches rely on activity-related engineered features to make the task of identifying activities possible. As such, the effectiveness of these approaches relies heavily on the nature of features, and their performance could change drastically if these features are not tuned well, or capture information characterizing the surrounding environment, leading to overfitting issues. Neural networks are used in the DL approaches. In this case, the detailed patterns needed to perform the classification of activities are learned automatically by the network, without requiring human engineering. However, in both cases, there is still a room for improvement.

DL has recently benefited from the advances in both the hardware and software sides (i.e., powerful Graphical Processing Units -GPUs- and libraries such as Tensorflow and PyTorch) making training large neural networks feasible in reasonable amounts of time. With relation to the IR-based AD systems, there are many constraints in the temperature, room coverage, and other environment-related limitations, all of which have an impact on the AD. Many techniques, such as background extraction [57–60], height estimation [61,62], people identification [50,63–66], and so on, have been used to improve these AD systems. However, such techniques do not

offer solutions for problems such as the noise, blurriness or distortion in the images. Nonetheless, most of the time, a model trained in one environment tend to drop in performance when tested in another environment with different characteristics such as the room temperature, the presence of heat-emitting devices, etc. making such systems less robust to environmental changes.

1.2 Existing Approaches

Many infrared array sensor-based AD have been proposed in the past few decades [15] [55] [67] [68] [69]. However, each of these AD systems has its own limitations. Most AD systems are based on conventional machine learning methods such as SVM, k-NN, etc. These conventional methods extract activity features manually to identify activities. As a result, the identification of activity with different people is less accurate. Mashiyama *et al.* [15] proposed an activity and fall detection technique [55] with an IR array sensor (8×8 pixels) mounted at the ceiling using the SVM and k-NN classifiers. This approach does not perform well on detecting certain activities, such as sitting, etc. Kobayashi *et al.* [67] proposed an AD system with two IR array sensors, one on the ceiling and the other on the wall and classified the activities using SVM. This approach was intended to improve on the previous one [55] by integrating the data obtained from both sensors. They achieved over 90% in the detection of all the activities. In particular, the detection of sitting activity increased from 78% to 93%. However, the detection of some other activities, including walking and falling, underperformed. Recently, Xiyui *et al.* [68] and Taramasco *et al.* [69] proposed to detect activities using IR array sensors placed on the ceiling and on the wall, respectively. Classifying activities such as walking, sitting, standing, etc. using Recurrent Neural Network (RNN) models achieved 85% and 93% accuracy, respectively. Furthermore, state-of-the-art methods on AD using various approaches, and their limitations are reviewed in Chapter 2.

1.3 Proposed Approaches

Based on the limitations of the existing systems, we strongly believe that the AD could be further improved, and more accurate systems could be built.

1. Therefore, first we propose an AD systems to improve the coverage and solve the temperature distribution problem using two IR array sensors and classified the activity using hybrid DL model by combining Convolution Neural Network (CNN) and Long Short-Term Memory (LSTM).

2. To further improve detection, we propose the AD systems using computer vision (CV) techniques to remove the noise in the image using Deep Image Prior (DIP) denoising technique, improve the quality of the image using Super-Resolution (SR), and improve the neural network's training using the data augmentation method.

1.3.1 AD Systems Using Dual IR Sensors

AD systems using dual IR array sensors approach use hybrid DL technique to detect the activities. One is placed on the wall, and another is placed on the ceiling. Both the sensors collect the data at eight frames per second. The consecutive frames collected by the sensors are classified using the hybrid DL model, regardless of the pattern of various temperature distribution pixels within them. The classification is performed on individual frames by CNN, and continuous sequences of frames by combining CNN and LSTM models with short time window size (5 frames which is less than a second). In addition, combine the ceiling data and wall data and classify each pair of frames using CNN and LSTM. This leads to an improvement of the classification accuracy of various activities thanks to combining both sensor data.

1.3.2 AD Systems Using Single IR Sensor

AD systems using a wide-angle IR array sensor with advanced DL based CV techniques. An IR array sensor is placed on the ceiling and collect data with various

resolutions (i.e., 24×32, 12×16, and 6×8). And apply the advanced DL techniques of SR and denoising to enhance the quality of the images. Then we classify the images/sequences of images using a hybrid DL model combining a CNN and a LSTM. We use data augmentation to improve the training of the neural networks by incorporating a wider variety of samples. The process of data augmentation is performed by a Conditional Generative Adversarial Network (CGAN). By enhancing the images using SR, removing the noise, and adding more training samples via data augmentation, to improve the classification accuracy of the neural network. On employing these DL techniques to noisy IR images leads to a noticeable improvement in AD performance.

1.4 Contributions

We propose a lightweight Deep Learning model for activity classification that is robust to environmental changes. Being lightweight, such a model can run on devices with very low computation capabilities, making it a base for a cheap solution for activity detection.

The activities are performed in all possible positions within the sensor coverage area irrespective of the sensor position. Most of the existing works require the subjects to perform the activities only in front of or under the sensor. In such a case, the blurriness and noise is due to the sensor characteristics the imprecision of the sensor capturing the temperature in a stationary position of the same activity, lead to a noticeable drop in performance. Our proposed neural network architecture manages to address this issue by exploiting the temporal changes in the frames to identify the activities accurately.

We identify the activity using a time window of less than 1 second. Despite the smaller time window, we have remarkably enhanced the classification accuracy in comparison to conventional works, which require a larger time window.

Low Resolution (LR) sensors are always preferred over High Resolution (HR) ones if they provide similar performance. This is thanks to their lower risk of privacy

invasion and cheaper cost. We demonstrate that it is possible to use the LR data to achieve classification performance that is nearly identical to that of the classification of the HR data, namely 24×32, by using deep learning techniques such as Super-resolution, denoising, and CGAN.

1.5 Outline of Dissertation

The remainder of this thesis is organized as follows.

- Chapter 2 discusses the various AD approaches that use IR array sensors, showcasing their limitations and motivations of the thesis.
- Chapter 3 elaborates the proposed AD approach with dual IR array sensors based on the corresponding framework, with performance evaluation through experiments.
- Chapter 4 further elaborates the proposed AD approach using advance DL based CV techniques with corresponding framework, and performance evaluation by experiments.
- Chapter 5 concludes this thesis and indicates future research directions.

For better understanding this dissertation, the organization and the relation among key chapters are shown in Figs. 1-3 and 1-4, respectively. Also, Tables 1.1 and 1.2 list the limitations of existing approaches of AD using IR array sensor, and the contributions of Chapters 3 and 4, respectively.

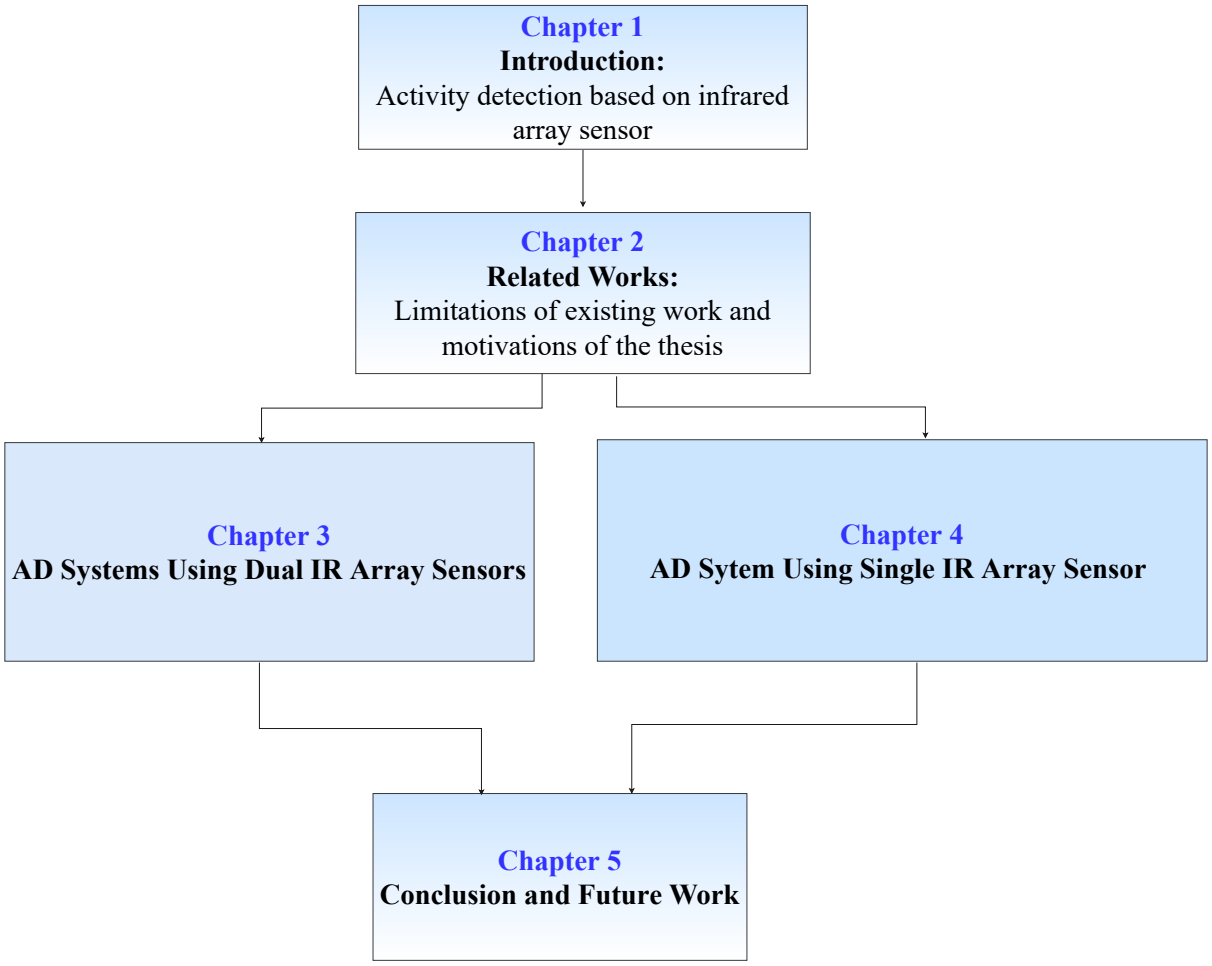


Figure 1-3: The organization of this thesis.

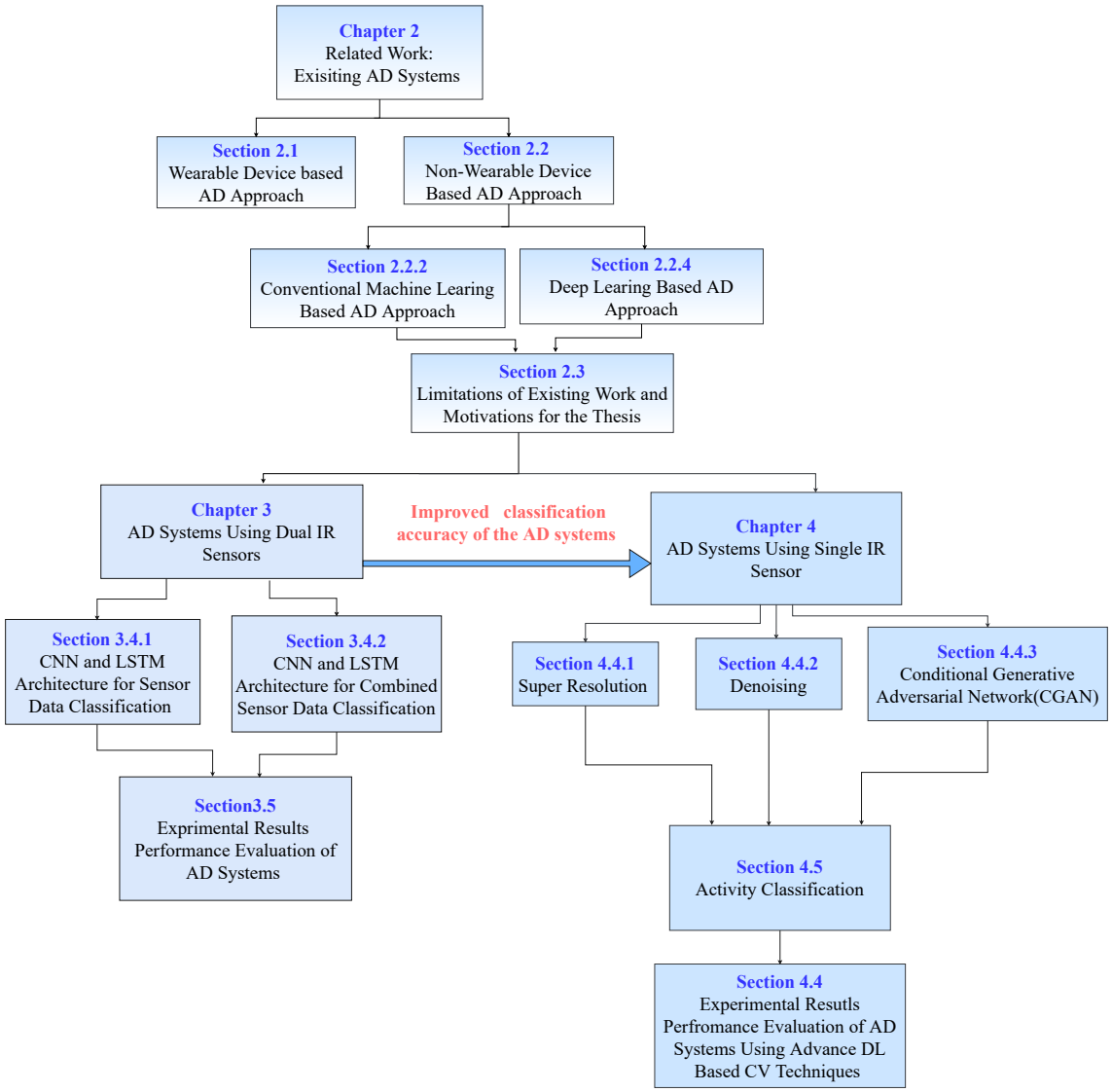


Figure 1-4: The relationship among the key chapters.

Table 1.1: Limitations of conventional machine learning based AD approaches and the contributions of chapter 3.

Research Problem	• Developing the effective AD using the wide angle IR array sensor.
Limitations of Existing Approaches	• Conventional machine learning approaches [55] [15] [67] relies heavily on manually engineered features. • In [68] [55] [15] uses a long time window size equal to 20 frames to identify the activity. However, despite the relatively long time window, the performance of these method still needs to be improved.
Proposed Approach	• AD approach that also uses LR IR array sensors placed on the ceiling and the wall. • The hybrid DL model combining CNN and LSTM. In this model, activity-related features are automatically created and learned by neural network. • Our model uses a window size equal to 5 frames (i.e., less than a second) to identify the activities.
Improvements	• The performance of AD is improved in wide coverage area by combining ceiling and wall sensor data. • Despite using a shorter time window, our hybrid DL model outperforms the conventional approaches.

Table 1.2: Limitations of existing AD approaches and contributions of chapter 4.

Research Problem	• Developing the AD systems using advance DL based CV techniques using wide angle IR array sensor.
Limitations of Existing Approaches	• The raw data from the IR array sensor is noisy and it affects the AD [70]. • Most of IR based AD use multiple sensors to improve the performance • Due to multiple sensors deployment. [71] [72], cost is high and also it is not robust to the environment.
Proposed Approach	• Our AD systems use advance DL based CV techniques to remove the noise and blurriness of the image. • We use advance data augmentation method to improve the robustness to the environment. • We use quantization to optimize the neural network for low-power devices, reducing deployment costs.
Improvements	• Using a single IR array sensor, we achieved LR data classification accuracy equal to HR data classification accuracy. • The optimized neural network can run on low powered device with advance DL methods.

Chapter 2

Related Work

2.1 Wearable Device Based AD Approach

2.1.1 Various Wearable Devices

Activity Detection using Accelerometer

A device that measures the acceleration or rate of change of velocity of a body in its instantaneous rest frame is known as an accelerometer. There are single-axis and multi-axis accelerometer models available to detect the magnitude and direction of proper acceleration as a vector quantity. Accelerometers are commonly used to detect orientation, vibration, shock, and falling in a resistive environment [73, 74]. Micro-electromechanical systems (MEMS) accelerometers are found in almost all modern portable devices. These accelerometers detect screen orientation, position, and so on. They are, however, perfectly capable of detecting fall events [75–78]. For fall detection, data from accelerometers can be used in machine learning, statistical models [76], or threshold-based algorithms [75].

Activity Detection Using Smartwatch

Smartwatches have created new opportunities for smart health [79–81], and wearable devices can detect and record a user's health-related daily activities unobtrusively. We can collect motion related sensor data from the smartwatch to detect daily activities

through analysis of wrist movement patterns because it is integrated with motion sensors (e.g., accelerometer, gyroscope, etc.) and worn by a user on his/her wrist most of the time. There are numerous applications for smartwatch-based activity recognition. It can function more intelligently if it is aware of what its user is doing. Physical inactivity has been used primarily to improve people's health and wellbeing. By providing accurate, real-time information about sedentary behavior and exercise, activity monitoring can help combat both inactivity and overeating. Because it is always readily and unobtrusively accessible, the smartwatch is perfectly positioned to convey this information, which is one of the reasons why smartwatch manufacturers tout its potential to improve health. Despite the fact that smartwatches have several activity recognition applications [8, 82–87]

2.1.2 Existing Works

Dinis *et al.* [7] proposed a human activity recognition approach using the smartphone inertial sensors. In their approach, the user has to carry a device (i.e., the phone in this case), and data are obtained from the device. Data for activities both indoors and outdoors are collected. The data are preprocessed and classified using ConvolutionalLSTM (ConvLSTM). The overall classification accuracy reaches 73%.

Andrea *et al.* [88] proposed a method for classifying various physical activities using a wearable accelerometer sensor. The data acquired by the accelerometer sensor are used to derive features based on the linear acceleration component due to the body motion and the gravitational acceleration component. Following feature extraction, a classification task is carried out using a variety of probabilistic and geometric approaches. The classifier with the best performance of classification accuracy 92.2% was based on a Hidden Markov Model (HMM).

Davide *et al.* [89] proposed an approach that uses smartphones for AD. In this approach, subjects must always hold the smartphone. Human activity signals are obtained from smartphone inertial sensors. Features are extracted manually from the signals received by the sensors and classify the activities using SVM with an overall accuracy of 87%.

Balli *et al.* [90] proposed a smartwatch-based approach for human activity detection. The device must be worn by the subject and data are obtained from the sensor. Using various machine learning methods such as SVM, k-NN, and random forest algorithm, features are extracted from the collected data. This study shows that the random forest algorithm performs better than SVM and k-NN.

Wearable device-based methods have their performance varying quite widely based on the type of sensor and the machine learning algorithm used. Nevertheless, they have their own limitations. A common shortcoming among them, however, is the need for manual extraction of features. Furthermore, the inconvenience of carrying such devices continuously is a drawback inherent to wearable devices, and cannot be avoided.

2.2 Non-Wearable Device Based AD Approach

In wearable device, there is a risk of device damage if they fall accidentally. With that in mind, non-wearable devices have several advantages over wearable ones, such as avoiding any physical contact with the person, which reduces the burden on the elderly. The non-wearable device-based systems such as ones based on cameras, sensors [15], antennas [11], Light Detection and Ranging (LiDAR) [13], Wi-Fi [14], and other similar devices require these devices to be strategically placed in specific locations to monitor the elderly's activities. That being said, non-wearable devices have several significant disadvantages, including privacy concerns, coverage issues, and so on.

Despite the limitations such as the privacy issues arising from the use of cameras, the coverage issues arising from the use of radars, and the compatibility issues raising from the use of wireless sensors, non-wearable devices are more convenient for the elderly. With the recent introduction of the IR array sensor, these issues have been mainly addressed. In indoor environments, these sensors are less intrusive and easier to use. The IR array sensor detects the heat generated by the human body and projects it onto a low-resolution matrix, which can then be treated as an image.

The non-intrusion from a privacy standpoint, the easy positioning/set-up, and the better coverage resulting in a wider detection area are just some of the advantages of this technology. Furthermore, its low cost makes it practical to deploy for real-world usage.

2.2.1 Various Non-Wearable Devices

Activity Detection Using WiFi

Electromagnetic signals in the radio or microwave spectrum comprise the wireless medium. These signals contain binary data. The channel data resulting from human interference with the wireless medium can be utilized for activity detection by machine learning or statistical models [91,92]. WiFall was developed by Wang et al. [92]. Human activities have an impact on the wireless medium. WiFall interprets the time variability and unique diversity of Channel State Information (CSI) as evidence of human activity. As almost all of the current wireless infrastructure already includes CSI, WiFall doesn't need wearable tech, special environmental changes, or even hardware modifications. WiFall was used on laptops with commercial 802.11n NICs in the system. CSI can estimate the channel properties of a communication link. Human motion can also be detected because it affects wireless propagation space, causing different patterns in the received signal. Based on the features extracted from the anomaly patterns, a one-class SVM was used to distinguish human falls. Laboratory experiments with WiFall yielded an 87% success rate and an 18% failure rate.

Activity Detection Using Radar

Radars are devices that use radio waves to track objects and determine their position, size, and velocity. A radar system typically includes a transmitter capable of producing electromagnetic waves in the radio and microwave spectrums, a receiving antenna, a receiver, a transmitting antenna, and a processor to determine the characteristics of the objects. Radio waves are transmitted by the transmitter. The objects cause these waves to reflect. Reflected waves can be used to determine the object's position and

speed [93]. Fall detection systems have made extensive use of Doppler radars [94, 95]. Using the Doppler effect, Doppler radars can detect moving targets at a great distance. Doppler radars send out microwave signals and measure how the frequencies of the returned signal change as the objects move. In order to detect falls from radar data, various signal processing techniques are typically utilized [94, 96–100].

Activity Detection Using LiDAR

Lidar is a type of sensing technology that determines the distance between a sensor and a target. The Lidar can estimate the distance to an person by emitting light and measuring the time it takes for the light to reflect from person and reach the receiver. In technical terms, the emitter, also known as the waveform generator, generates a laser wave-form (or an array of laser waveforms). The laser waveform travels through a medium until it hits a target. The Lidar light bounces off the detected object and returns through the same medium, where it is captured by the receiver. Depending on the medium being explored (air, water, vacuum, etc.), the precision, and the penetration required, different wave-lengths are used. In an application such as activity detection [101, 102], the obvious medium is air, which has a light speed close to that of a vacuum. A 2-D Lidar is simply a Lidar with a rotating motor that allows it to cover 360 degrees. The Lidar sensors emit light, receive light, estimate the distance to objects at various angles (points), and map them to a 2-D map by rotating. While rotating, the Lidar emits light from its emitter. The receiver captures the reflected light, and because the speed of light is constant, the distance to the object at that particular angle can be calculated. By taking several measurements while rotating, the lidar can create a 2-D map that can be used to identify people and their activities in the vicinity of the Lidar.

Activity Detection Using IR Sensor

An infrared sensor is a type of electronic sensor that detects infrared light emitted by objects in its field of view. Although infrared sensors can detect general movement, they cannot provide information about the moving subject. Because humans emit

Table 2.1: Comparison of various AD devices.

Parameters	IR array sensor	LiDAR [13]	Radar [107]	Wi-Fi [14]
Device cost	60$	500$	410	100$
Power source	30$	100$	40$	50$
Embedded computer	40$	700$	100$	40$
Total cost	130$	1300$	560$	190$
Size (L×W×H mm)	56.5×85.6×17	76×76×41	171×158×41	212×183×34
Inference time	0.43 ms	126 ms	5.7 ms	7.9ms

mostly infrared radiation, IR sensors can be used to track human movement [103,104]. Infrared-based systems are primarily used for surveillance. IR sensor data is typically used to generate 3D images or blocks that represent environmental infrared radiation information [105]. Following feature extraction, a variety of machine learning or statistical models are used to detect falls and ADL events [105, 106].

In Table 2.1, we show a comparison of the various devices (thus the respective approaches which employed them) in terms the cost, computational time and size of the devices.

2.2.2 Conventional Machine Learning Based AD Approach

Support Vector Machine (SVM)

SVM, or Support Vector Machine, is a linear model that can be used to solve classification and regression problems. Non-linear mapping of input quantities to a very high-dimensional feature space. A line decision surface is built in this feature space [108]. According to [109], SVM transforms the original training data into a higher dimension using a non-linear mapping. Within this new dimension, the optimal separation hyperplane is sought. SVMs can be used for both numerical prediction and classification. They have been used in a variety of applications such as activity detection [110–114], object detection [115–118], and so on.

Naïve Bayesian Classifier

It is a unique subset of machine learning algorithms that deals with the classification task [119]. The "Bayes theorem" [120] serves as the foundation for this. This algorithm assumes the prediction variables are independent [121]. In other words, the presence of a set of characteristics [122] in one data set does not imply the absence of another character in another data set.

Random Forest

Random Forest [123] is a popular feature selection algorithm that automatically calculates the importance of each feature without the need for additional programming. This allows us to select a more limited set of features. The random forest has the following advantages [124]: high precision, the introduction of randomness makes randomforests avoid overfitting problem, it have good denoise ability (can handle outliers better), can handle very high-dimensional data without feature selection, it can handle both discrete data and continuous data, and high training speed. The random forest has some drawbacks [125], including poor interpretability when there are lots of decision trees present, which increases training time and space requirements.

2.2.3 Existing Works

Device-free approaches for AD have attracted more attention over the last few years. Most of them still rely on conventional machine learning techniques. The conventional approach use manually engineered features to identify activities. In [15] and [55], Mashiyama et al. proposed two similar approaches for fall detection and AD, respectively. In their work, they used a single IR array sensor (8×8 pixels) attached to the ceiling. Data are collected from the sensor with a fixed time window size of frames. From the data collected in each scenario, they extracted four features which they use to run a classification task using conventional machine learning to identify the activity that the subject is performing. The features are extracted from the consecutive frames where the motion is detected. The features used are: the maximum number of

pixels that changed during these consecutive frames, the maximum of the variance of temperature among the pixels, and the maximum temperature difference in the same pixel before and after the activity. In their work, the authors ran several experiments. However, the best results obtained are as follows. Using a k-NN classifier, their approach achieved an accuracy for fall detection equal to 94% classification accuracy. Using a SVM classifier, the accuracy achieved for AD is 100% for no event, 94.8% for stopping, 99% for walking, and 78% for sitting. Despite its merits, this system has a few shortcomings. To begin with, no processing is done on the data. Nevertheless, due to the LR data used (8×8 pixels), the noise is extremely high, making the values for features extracted present a high level of error. As a result, the classification performance was affected significantly for some activities.

Kobayashi *et al.* [67] proposed an AD system based on the temperature distribution captured by two IR array sensors placed on two different locations (i.e., on the ceiling and on the wall) that collect the data simultaneously. An SVM classifier is used to run a classification task on the data collected by both sensors together. Their proposed activity recognition system addressed the limitations of the previous work [15], namely the poor classification accuracy for some activities. For instance, the detection of the sitting activity has been improved from 78% to 93%. This system's overall classification accuracy is above 90% for all the activities including walking, falling, sitting, etc. While this work addressed some of the limitations of the previous one by introducing a few features and using two sensors simultaneously, it still has a few shortcomings. For instance, no noise removal is done in this approach that may cause the performance degradation. The improved performance is mainly attributed to the combination of two sensors' data. Nevertheless, this system has been tested in a single environment.

Taniguchi *et al.* [69] has proposed a fall detection system using two thermal array sensors (16×16 pixels). One is placed on the wall, and the other on the ceiling. All the activities are carried out under and in front of the sensors. Both the sensor data are combined, and the temperature distribution is used to distinguish fall activities from non-fall activities. Their approach achieved an accuracy equal to 72%. This system,

however, relies on one of the oldest time series analysis approaches like time-series posture transition diagram, and the sum of temperature distribution. Several newer machine learning models perform much better.

2.2.4 Deep Learning Based AD Approach

The deep learning techniques utilize neural networks. In this case, the neural network learns the detailed patterns required to perform activity classification automatically, without the need for manually engineered features.

Deep Neural Network

Deep neural networks (DNN) are derived from artificial neural networks (ANN) (ANN). Deep neural networks (DNN) have more hidden layers than traditional neural networks (ANN) (deep). DNN can learn from large amounts of data with more layers. DNN is typically used as the dense layer in other deep models. In a convolution neural network, for example, several dense layers are frequently added after the convolution layers.

Hand-engineered features were first extracted from the sensors [126] before being fed into a Deep Neural Network model. Similarly, [127] In formed PCA before employing DNN. Because DNN was only used as a classification model after hand-crafted feature extraction in those studies, they may not generalize well. And the network was a little shaky. In [128] improved performance by using a 5-hidden-layer DNN for automatic feature learning and classification. Those studies found that when the HAR data is multidimensional and the activities are more complex, adding more hidden layers can help the model train well because their representation capability is stronger [129]. In certain circumstances, however, it is necessary to consider additional information to help the model achieve a more accurate fit.

Convolution Neural Network

Convolutional Neural Networks (ConvNets, or CNN) make use of three key concepts: sparse interactions, parameter sharing, and equivariant representations [130]. Following convolution, there are typically pooling and fully-connected layers that perform classification or regression. CNN is capable of extracting features from signals and has shown promise in image classification, speech recognition, and text analysis. CNN has two advantages over other models when applied to time series classification, such as HAR: local dependency and scale invariance. Local dependency in HAR refers to the likelihood of nearby signals being correlated, whereas scale invariance refers to the scale-invariant for different paces or frequencies. Because of CNN's effectiveness, the majority of the work surveyed was in this field.

When applying CNN to HAR, several factors must be considered, including input adaptation, pooling, and weight-sharing.

1. Input modification. In contrast to images, most HAR sensors generate time series readings such as acceleration signals, which are temporal multidimensional 1D readings. Before applying CNN to those inputs, input adaptation is required. The main idea is to modify the inputs to create a virtual image. There are two types of adaptation: model-driven adaptation and data-driven adaptation.

    (a) The data-driven approach considers each dimension to be a channel and then performs 1D convolution on it. The outputs of each channel are flattened to unified DNN layers after convolution and pooling. In [131] an early work in which each dimension of the accelerometer was treated as one channel, similar to RGB of an image, and then convolution and pooling were performed separately. In [132] also proposed using 1D convolution in the same temporal window to unify and share weights in multi-sensor CNN. In addition to this line, [133] resized the convolu- tion kernel to get the best kernel for HAR data. [128, 134, 135] are examples of similar work. This data-driven approach, which is simple and easy to implement, treats the 1D sensor reading as a 1D image. This approach has the disadvantage

of ignoring the interdependencies between dimensions and sensors, which may affect performance.

(b) The model-driven approach resizes the inputs to a virtual 2D image so that a 2D convolution can be used. This approach is typically used for non-trivial input tuning techniques. While [136] created a more complex algorithm to convert the time series into an image, [137] combined all dimensions to create an image. In [138] used modality transformation to convert pressure sensor data to an image. Similar work can be found in [139,140]. This model-driven approach can make use of sensor temporal correlation. However, mapping time series to images is a difficult task that necessitates domain knowledge.

2. Pooling:Convolution-pooling is a common combination in CNN, with most approaches performing max or average pooling after convolution [134, 137, 141], [19,31,48]. In addition to preventing overfitting, data pooling can accelerate the training process for large datasets [129].

3. Weight distribution. Weight sharing [135, 142]is an effective way to accelerate the training process for a new task. In [131] used a relaxed partial weight sharing technique because the signal in different units behaved differently. In [143] investigated the performance of various weight-sharing techniques using a CNN-pf and CNN-pff structure. It has been demonstrated in the literature that partial weight-sharing can improve CNN performance.

Autoencoder

The hidden layers of the Autoencoder learn a latent representation [144] of the input values, which can be thought of as an encoding-decoding procedure. The autoencoder's goal is to learn more advanced feature representations using an unsupervised learning schema. The stack of some autoencoders is known as a stacked autoencoder (SAE). SAE considers each layer to be the basic model of an autoencoder. The learned features are stacked with labels to form a classifier after several rounds of training.

In [145,146] used SAE for HAR, adopting the greedy layer-wise pre-training [147] and then performing fine-tuning. In contrast to those studies, In [148] investigated the sparse autoencoder by including noise and KL divergence in the cost function, indicating that the addition of sparse constraints could enhance HAR performance. SAE has the advantage of being able to perform unsupervised feature learning for HAR, making it a powerful tool for feature extraction. However, SAE is overly reliant on its layers and activation functions, making it difficult to find the best solutions.

Restricted Boltzmann Machine

The restricted Boltzmann machine (RBM) [149] is a bipartite, fully connected, undirected graph with a visible and hidden layer [147]. By treating every two consecutive layers as an RBM, the stacked RBM is referred to as a deep belief network (DBN). Fully-connected layers are frequently added after DBN/RBM. In pre-training, most studies used Gaussian RBM in the first layer and binary RBM in the subsequent layers [150–152]. A multi-modal RBM was developed by [153] for multimodal sensors, in which an RBM is built for each modality of the sensor before the output from all the modalities is combined. To extract the important features, [139] added pooling after the fully-connected layers. In [154] used a contrastive gradient (CG) method to fine-tune the weight, which allows the network to search and converge quickly in all directions. In [155] went on to implement RBM on a mobile phone for offline training, demonstrating that RBM can be very lightweight. RBM/DBN, like autoencoder, can perform unsupervised feature learning for HAR.

Recurrent Neural Network

By utilizing the temporal correlations between neurons, recurrent neural networks (RNN) are widely used in speech recognition and natural language processing. LSTM (long-short term memory) cells are frequently combined with RNN, with LSTM serving as memory units via gradient descent. Few works [128, 156–158] used RNN for HAR tasks, where learning speed and resource consumption are the primary concerns for HAR. In [158] first investigated several model parameters before proposing

a relatively good model capable of performing HAR with high throughput. In [156] proposed a binarized-BLSTM-RNN model in which all hidden layer weight parameters, input, and output are all binary values. The main goal of RNN-based HAR models is to work well in environments with few resources while still getting good results.

Hybrid Model

A hybrid model is the result of the combination of several deep models [159]. The combination of CNN and RNN is one emerging hybrid model. In [132, 160] provided excellent examples of combining CNN and RNN. In [160] demonstrates that the performance of 'CNN + recurrent dense layers' outperforms 'CNN + dense layers'. In [138] also shows similar results. The reason for this is that CNN can capture spatial relationships while RNN can use temporal relationships. Combining CNN and RNN [161, 162] may improve the ability to recognize various activities with varying time spans and signal distributions. Other research combined CNN with models like SAE [163] and RBM [164]. CNN is used to extract features in these works, and generative models can aid in the training process. We anticipate more research in this area in the future.

2.2.5 Existing Work Based on IR Array Sensors

Xiyui *et al.* [68] has proposed a robust fall detection system using an infrared array sensor (8×8 pixels). The sensor is placed on the wall in this system. The different activities are carried out in parallel and perpendicular to the sensor. The data is pre-processed by applying a Gaussian filter, and a median filter then forwarded to an LSTM and a Gated Recurrent Units (GRU) recurrent neural networks to be classified. The system achieved an accuracy equal to 75% using LSTM and 85% using GRU. The activities in this system are performed in limited positions, and the accuracy of the classification is low in both the algorithms.

Taramasco *et al.* [165] has proposed a fall detection system using an infrared ar-

ray sensor (1×16 pixels). The sensor is placed on the ceiling, and the subjects have carried out a variety of activities, the data of which are collected using the sensor. Activities are classified using a RNN, which is used for classifying sequences with different architectures, such as LSTM, GRU, and Bi-LSTM (Bi-directional LSTM). Their performance varies. However, Bi-LSTM performed the best, achieving an accuracy equal to 93%. The Bi-LSTM approach requires a high computation device to run, limiting its usability on low computation devices.

Javier *et al.* [166] proposed an approach for fall detection that relies on a single IR array sensor with a 32×31 resolution installed on the ceiling. In this work, the authors use conventional data augmentation techniques such as rotating and cropping the image to improve the classification. The classification is done by three different types of CNN, the best of which reaches an accuracy equal to 92%.

Matthew *et al.* [71] proposed an unobtrusive pose recognition using five IR array sensors with 32×31 resolution. In their work, the authors proposed to install a single sensor on the ceiling and the other four ones on four corners of the room. The data are collected and classified using a CNN. In their work, the authors analyzed the performance of classification of data collected by the individual sensors, as well as their combination. They achieved an overall F1-score equal to 92%. An interesting finding of theirs is that the ceiling sensor data classification performance is poor comparatively. This work did not perform the classification taking into account the temporal changes in the collected frames due to activities. Nevertheless, they used five sensors, which makes it relatively an expensive solutions to justify a marginal improvement in performance.

Cankun *et al.* [70] proposed a multi-occupancy fall detection system using an IR array sensor with a 32×31 resolution. This work decomposes multi-occupancy data from the sensor using image binarization, contours detection, and single occupancy sub-image. The features are extracted and classified using CNN. The highest average classification accuracy achieved in this work is 98.39%. However, a few true negatives were recorded for the class "fallen", which presents a major drawback as this class is the most crucial to detect. The misclassification is mainly due to the images being

blurry and with high contrast, as well as to the person falling at the edge of coverage of the sensor.

Tianfu *et al.* [167] proposed human action recognition using two IR array sensors, whose resolution is 24×32. One of them is placed on the ceiling and the other is placed on the wall. Sequentially, the two sensor data go through a set of pre-processing operations. These include a quantification, a time-domain filtering, and a removal of the background. The pre-processed data are used to locate the human target and detect the activity (s)he is performing. The classification of the activities is done by a CNN. The highest classification accuracy obtained is 96.73%. Although multiple sensors are used in this work, this method failed to detect the position of the human when (s)he is near the edges of coverage of the sensor. This is mainly due to the blurriness and noise in the images.

Miguel *et al.* [72] proposed a fall detection system using two IR array sensors: one with HR of 80×60 placed on the ceiling and the other with a LR of 32×31 placed on the wall. The collected thermal data are in a fuzzy representation, and the activities are classified using three different CNNs and their respective results are compared. Nonetheless, in their work, the authors used a traditional data augmentation method to improve the classification accuracy by rotating and cropping the images. Interestingly, thanks to its wide angle lens allowing for a wider coverage, the LR sensor data classification performs better than HR one. The highest classification accuracy they obtained is 94.3%. In this work, the authors did not combine the data nor did they perform the sequence data classification. Nonetheless, they did not consider removing the noise or enhancing the resolution of the images.

Tateno *et al.* [168] proposed a fall detection system using one IR array sensor placed on the ceiling. The data is preprocessed by applying noise removal and background subtraction, upon which the target in motion is located and his/her activity is classified using a 3D-CNN and an LSTM, separately. In this work, the authors aimed to detect the activity of multiple people. Using cross validation data, the highest classification accuracy reached 98.8% of 3D-CNN and 94.9% of LSTM. Despite its high accuracy, this approach has not been proven to be robust. As the authors did

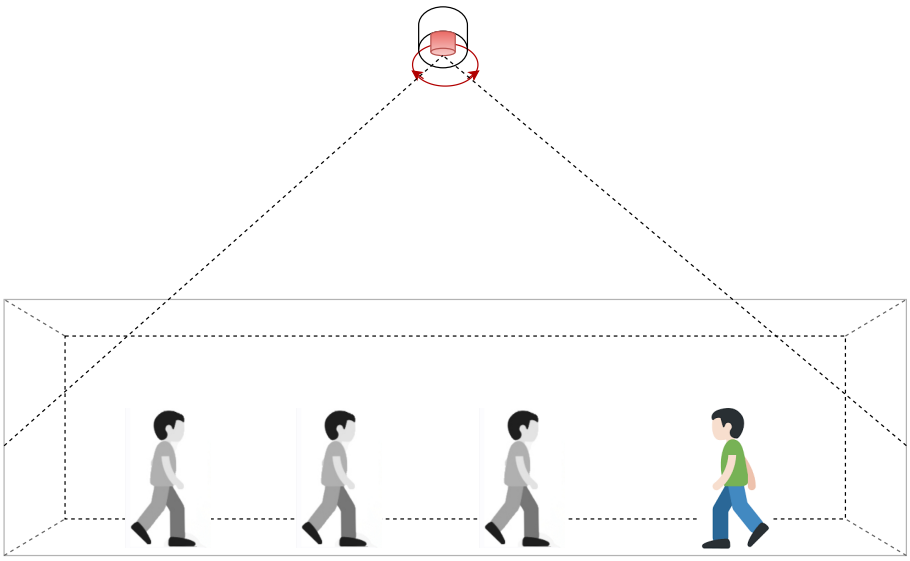


Figure 2-1: The sensor placed at the ceiling

not run a classification on unknown test data taken in a different environment. This leads us to believe that chances of having similar performance in real-world application on new unseen environments are not to be expected. Nevertheless, the authors used Gaussian filter for noise removal, which is a linear smoothing filter that results in information loss.

2.2.6 Position of IR sensor at Various Locations

Sensor Placed at Ceiling:

The ceiling sensor positioning has both merits and demerits [169–173]. When we choose to place the sensor on the ceiling, it needs to be placed at a sufficient height in order to cover a large area shown in the Fig 2-1. However, in the case of an infrared sensor, as the subject moves farther away from the sensor the heat distribution in the image reduces making it harder to identify the activity. Thus this positioning of the sensor can cover a large area but the quality of the data collected is affected thereby making it inadequate.

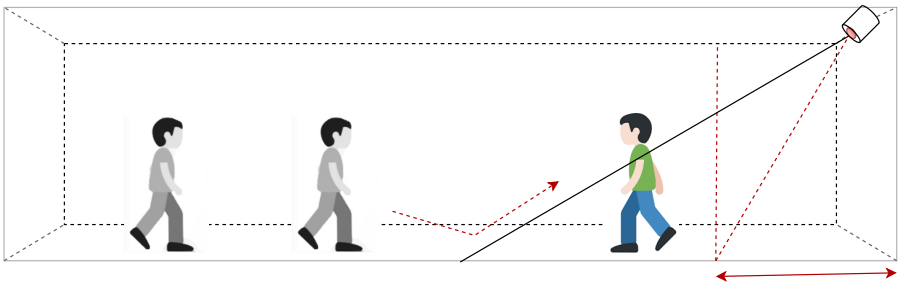


Figure 2-2: The sensor placed at the corner of the room

Sensor Placed at the Corner of the Room:

When the same sensor is placed at one corner [174–178] of the room shown in the Fig. 2-2, the coverage is affected because of the angle of inclination. This results in a better view at the bottom and the top view being affected negatively. Thus when the person moves farther away from the sensor he will be out of the coverage area gradually, and this will make it harder to detect the activity towards the edge of the coverage area.

Sensors Placed Linearly at the Ceiling:

When multiple IR array sensors are placed continuously in the ceiling [179–182] show in Fig. 2-3, it will enable coverage of a larger area with lower height. Thus activity detection is improved further. However, the job of combining multiple sensor data [183] and ordering them to find the corresponding activity is a difficult task since the features for the activities vary on account of the overlap between the field of view of the sensors. Moreover, this kind of activity detection system only works in a specific environment and is not robust to work in other environments.

Sensors Placed at Bottom Four Corner of the Room:

In the case where the sensor is placed at the bottom corner [71, 184, 185] edge of the room shown in Fig. 2-4, the peripheral view of the sensor struggles to register the person's activity and is hard to detect. An attempt to rectify this issue by placing

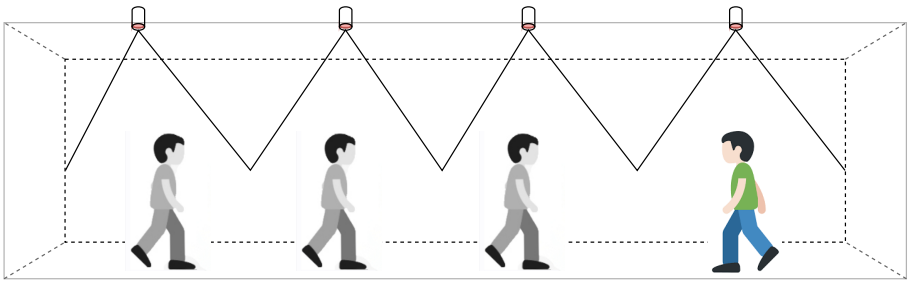


Figure 2-3: The multiple sensors deployed linearly

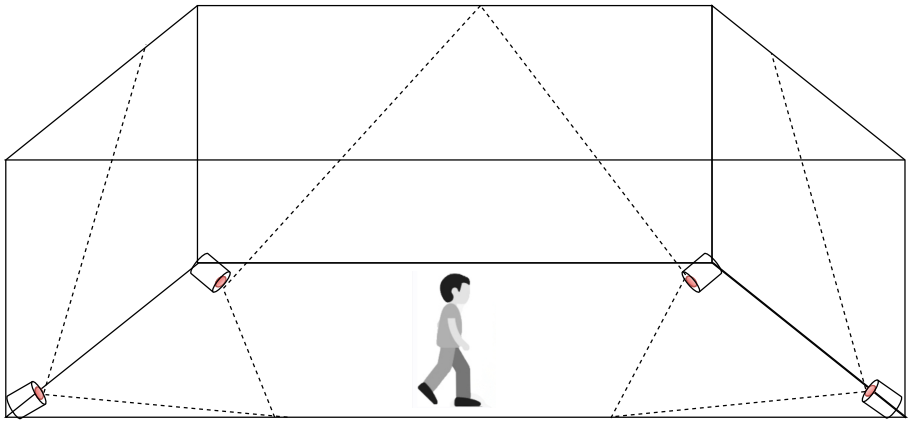


Figure 2-4: The sensors placed at the bottom corner of the room.

multiple sensors in all four bottom corners of the room causes difficulty in ordering of the data to identify the corresponding activity.

2.3 Limitations of Existing Work and Motivations for the Thesis

2.3.1 AD Using Hybrid Deep Learning

Based on the limitations of the existing approaches, we have chosen to use a two 32×24 pixels resolution infrared array sensors. One is placed on the ceiling and another on the wall. The sensor's resolution is higher than that of the existing 8×8, 1×16, 16×16 pixels resolution sensors. Nonetheless, the sensor that we are using has a wide-angle allowing for a coverage area that is much wider than that of the

other sensors. The activities are performed in all possible positions within the sensor coverage area irrespective of the sensor position. Most of the existing works require the subjects to perform the activities only in front of or under the sensor. In such a case, the blurriness and noise due to the imprecision of the sensor capturing the temperature in a stationary position of the same activity. This makes this type of images non-appropriate for feature extraction using conventional machine learning methods. The features in these methods are extracted using temperature distribution changes. Clear images are easily classified using these conventional methods, whereas, different pattern of temperature distribution of same activities are much harder to classify. Deep learning techniques are thus better suited to classify the these kind of images. In the field of deep learning techniques, the activities' features are automatically learned by the neural network.

2.3.2 AD Using Computer Vision Techniques

Most of the state-of-the-art work related to the detection of activities relies heavily on a multitude of sensors and is restricted by the environment conditions. Such AD systems are less effective when deployed on new unseen environments. Nevertheless, these works, for the most part, ignore the effect of noise and blurriness produce due to the pattern of temperature distribution on their performance. This is a key point to address as the noise level in low-resolution IR images is relatively high. It has a significant impact on the detection of activity.

To develop the AD systems we choose one IR array sensor placed on the ceiling and conduct various experiments. It offers a number of advantages to protect the user's privacy, it also works in a variety of environments (in terms of luminance, including darkness). However, these sensors can only be used in a very small space due to their limited coverage.

That being said, it is fair to assume that LR sensors are always preferred over HR ones if they can provide similar performance. This is thanks to their cheaper cost and lower risk of privacy invasion. In our work, we aim to introduce a solution which uses LR sensors (i.e., 12×16 and 6×8) to perform AD with performance comparable to HR

ones. For the sake of our experiments, we chose to use a 24×32 resolution IR array sensor. We collect data in 12×16 and 6×8 resolutions, respectively. For these lower resolutions, we try to achieve a performance nearly equal to the one reported when using the 24×32 resolution. Using an LR image presenting a large amount of noise makes it hard to identify the activity the subjects are performing. Thus, reaching a performance comparable to that obtained using HR frames is a challenging task. Another major challenge is the temperature distribution pattern varies due to the use of a wide-angle lens. The same activity manifests differently depending on where it has been performed, thus creates a different pattern. With the proposed classification model, and thanks to the use of data augmentation, we aim to classify the activities more accurately even with the patter of temeprature distribution changes. We apply some advanced techniques of DL to achieve better performance.

Table 2.2 shows a comparison of the existing work and their limitations. Most of the existing work performs the activity in front of the sensor. They did not do the activities in different positions within the coverage areas of the sensor. The different angles in addition to the temperature distribution pattern varies, produce different patterns for the same activity, making it hard for these AD systems to detect the activity, affecting their performance. Also, the majority of the work did not address issues related to noise in the data collected, and it only works well in specific environments. Another common limitation is the high computational and deployment cost required for these approaches. In addition, many of the works use two or more sensors to achieve the high performance which, again, is expensive for real-world deployment.

Table 2.2: A summary of the existing works that use IR sensors for AD alongside with their shortcomings.

Study	Resolution	# sensors	Position	Methods	Accuracy	Limitations
Mashiyama *et al.* [15]	8×8	1	Ceiling	SVM	94%	Data are highly noisy due to their low resolution. Few activities in a specific area, no detection of transition between activities.
Mashiyama *et al.* [55]	8×8	1	Ceiling	k-NN	94%	Due to the noise in the data, feature extraction is less effective.
Kobayashi et al [67]	8×8	2	Ceiling, Wall	SVM	90%	No reprocessing is done. Data are noisy. Activities are performed in very specific positions.
Xiyui *et al.* [68]	8×8	1	Wall	LSTM, GRU	75% and 85%	Very limited positions: activities are performed only in parallel or perpendicular to the sensor.
Taniguchi *et al.* [69]	16×16	2	Ceiling, Wall	Time series analysis	72%	Low accuracy due to the use of an old approach.
Taramasco *et al.* [165]	1×16	2	Opposite corner of the room	LSTM, GRU, Bi-LSTM	93%	High computation cost.
Javier *et al.* [166]	32×32	1	Ceiling	CNN	92% & 85%	Noisy and blurry image. Difficult to detect activities in high temperature areas.
Matthew *et al.* [71]	32×31	5	Ceiling and all corners	CNN based on alexnet	F1-score 92%	Requires multiple sensors. Expensive to deploy in real-world.
Cankun *et al.* [70]	32×31	1	Ceiling	multi-occupancy fall detection MoT-LoGNN	98.39%	The misclassification of activities is mainly due to the image being blurry and having high contrast.
Tianfu *et al.* [167]	24×32	2	Ceiling, Wall	CNN	96.73%	Fails to detect the position of the human near the edges of coverage, due to the blurriness and noise in the images.
Miguel *et al.* [72]	32×31, 80×60	2	Ceiling, Wall	CNN	72%	No sequence data classification was performed. Noise removal and enhancement of the images were not performed.
Tateno *et al.* [168]	24×32	1	Ceiling	3D-CNN 3D-LSTM	93%	Gaussian filter is used to remove the noises, which causes a loss of information.

Chapter 3

AD Systems Using Dual IR Sensors

3.1 Introduction

There are a lot of non-wearable devices available for activity detection. They include but are not limited to radars, Wi-Fi, IR sensors, etc. Among these, we specifically choose to use the IR array sensor because it has several advantages. Not only does it protect the privacy of the user, but it also operates in a variety of environments (in terms of luminance, including darkness). Most of the applications and existing work relying on IR array sensors use ones with a resolution equal to 8×8 pixels. However, these sensors have very limited coverage and can be used only in a very small room.

Given the limitations, we listed above in Chapter 2, we have chosen to use a 32×24 pixels resolution IR array sensor. The sensor's resolution is higher than that of the existing 8×8, 1×16, 16×16 pixels resolution sensors. Nonetheless, the sensor that we are using has a wide-angle allowing for a coverage area that is much wider than that of the other sensors. The activities are performed in all possible positions within the sensor coverage area irrespective of the sensor position. Most of the existing works require the subjects to perform the activities only in front of or under the sensor. In such a case, the blurriness and noise is due to the sensor characteristics the imprecision of the sensor capturing the temperature in a stationary position of the same activity. This makes this type of images non-appropriate for feature extraction using conventional machine learning methods. The features in these methods are

extracted using temperature distribution changes. Clear images are easily classified using these conventional methods, whereas, different pattern of temperature distribution pixel images are much harder to classify. This makes DL techniques more suitable for images of this kind. In the field of DL techniques, the activities' features are automatically learned by the neural network. In this AD systems we use a hybrid deep-learning model to classify the same activity with different pattern of temperature distribution; the neural network automatically learns pattern activity features. Two sensors are used in the proposed system. One of the sensors is placed on the wall, and another one is placed on the ceiling. Both the sensors collect the data at eight frames per second and start simultaneously. After collecting the data, the proposed activity detection technique involves two stages. First, we classify the individual frames collected by the wall sensor and the ceiling sensor separately using a CNN. In the second stage, the output of the CNN is passed through a LSTM with a window size equal to 5 frames to classify the sequence of activities. Afterwards, we combine the ceiling data and wall data and classify each pair of frames using CNN. The output of the CNN is passed through the LSTM with a window size equal to 5 frames. This leads to an improvement of the classification accuracy of various activities thanks to combining both sensor data.

3.2 Framework of AD Systems

The overall framework of our proposed system is shown in Fig. 4-1. We collect data from our experiments, and then using CNN and LSTM, we perform the classification of the various activities. First, we classify individual frames collected by the wall sensor and the ceiling sensor separately using the CNN. We then pass the output of the CNN through the LSTM for sequential activity classification and check the performance on both the ceiling sensor data and the wall sensor data separately. Second, we combine the wall sensor data and the ceiling sensor data. Using CNN, we classify the individual pairs of frames of the activities and analyze the performance. The output of the CNN is passed through the LSTM for sequential classification of

Figure 3-1: A flowchart of the proposed system.

the activities. The outputs of CNN and LSTM using wall sensor data are represented by CNN_w and $LSTM_w$, respectively. In the same way, the outputs of CNN and LSTM using ceiling data are represented by CNN_c and $LSTM_c$, respectively. The output of CNN and LSTM using the combined ceiling sensor data and wall sensor data is represented by CNN_{cw} and $LSTM_{cw}$, respectively. We use these notations to differentiate between the different models and to make it easy to compare their performance.

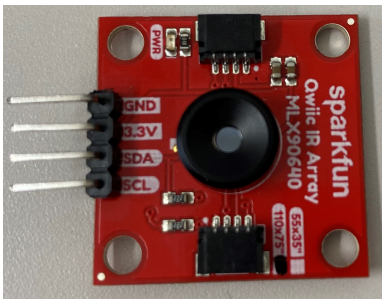


Figure 3-2: The wide angle IR array sensor used for our experiments.

Table 3.1: The technical specifications of the sensor.

IR sensor model: Qwiic IR Array	MLX90640
Camera	1
Voltage	3.3 V
Temperature range of targets	$-40°C \sim 85°C$
Absolute temperature accuracy	$\pm 2°C$
Number of pixels	768 (32×24)
Viewing angle	$110° \times 75°$
Frame rate	8 frames/second

3.3 Experiment Specification

3.3.1 Device Specification

We used two of the MLX90640 (Melexis corporation)[1] IR array sensor shown in Fig. 3-2. These sensors are capable of detecting heat rays from any thermal source. Table 3.1 displays the main sensor specifications. The sensor temperature range covers both the typical human temperature as well as indoor temperature. Nevertheless, the sensor can collect data at different frame rates. The sensor frame resolution is 32×24 pixels. The brighter the color is in the generated frames, the higher the temperature is.

The sensor is attached to a Raspberry Pi 3 model b+ as shown in Fig. 3-3. The Raspberry Pi is also equipped with a standard camera recording the same event

[1]https://www.melexis.com/en/product/MLX90640/

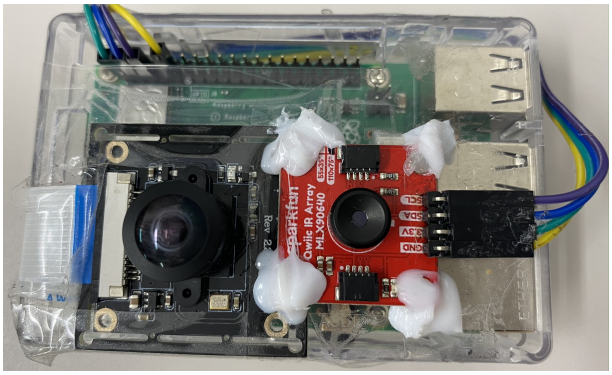


Figure 3-3: An image of the Rasberry Pi 3+ with the camera and the IR senors mounted which we used for collecting the data.

as the sensor. The data collected by the camera are used as ground truth and are used to annotate the sensor data. We prepared two sets of devices with the same configuration, one is placed on the wall, and the other is placed on the ceiling. The wall sensor and the ceiling sensor as well as their corresponding cameras collect data at the same rate of 8 frames per second (fps). The data are stored in the SD card mounted in the Raspberry Pi.

3.3.2 Environment Specification

The experiment has been set up in a large meeting room environment with a standard room temperature. Two IR array sensors are deployed in the room, one on the ceiling and the other on the wall. In Fig. 3-4, we show a simplified scheme of the sensor deployment and an example of a frame collected by the sensor.

Fig. 3-5 shows the coverage measurements according to the sensor specification. The sensor has a wide-angle: the coverage alongside the first angle is 110°, and alongside the other is 75°. Using these angles, we calculate the ceiling sensor coverage area, i.e., length × breadth, which we refer to as l_1 and l_2, respectively (which correspond to the coverage and the angles θ_1 and θ_2, respectively). The sensor is attached to the ceiling at a height equal to 2.60 m from the ground. We refer to this height as h_c. Based on the known values, the coverage area can be calculated using the following equations.

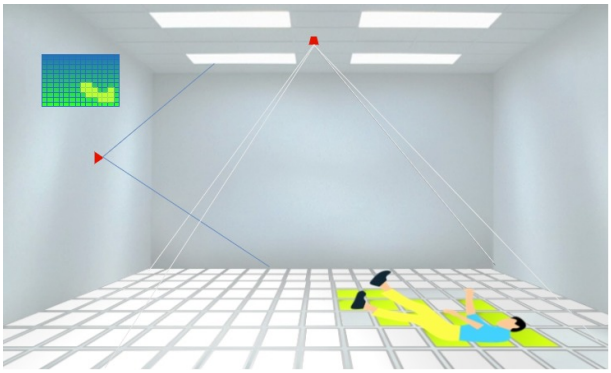


Figure 3-4: The experiment coverage area of the sensor.

$$l_1 = 2h_c \tan\left(\frac{\theta_1}{2}\right), \tag{3.1}$$

$$l_2 = 2h_c \tan\left(\frac{\theta_2}{2}\right). \tag{3.2}$$

The coverage at the ground level, however, is not realistic. In addition, in the case where the human is at the edge, barely his feet will be detected, as shown in Fig. 3-5. Therefore, we use α and β coefficients to ensure that coverage is sufficiently reliable. In consideration of our early experiments, α is set to be 0.81, and β is set to be 0.75. This will effectively cover an area whose length and breadth are equal to 7.40 m and 3.90 m, respectively.

The wall sensor is placed on the wall at a height 1.00 m from the floor, the height of the sensor represented in h_w. The coverage area of the wall sensor is shown in Fig. 3-5. We calculated the wall sensor coverage area using the angle of the sensor θ_1 and the h_w. Here, γ is the angle between the sensor coverage and the wall. The blind angle of the sensor, where the detection using the wall sensor is not possible, is represented by the distance d. Based on the known values of θ_1 and h_w, we calculated d using the following equations.

$$\gamma = 90^\circ - \left(\frac{\theta_1}{2}\right) \tag{3.3}$$

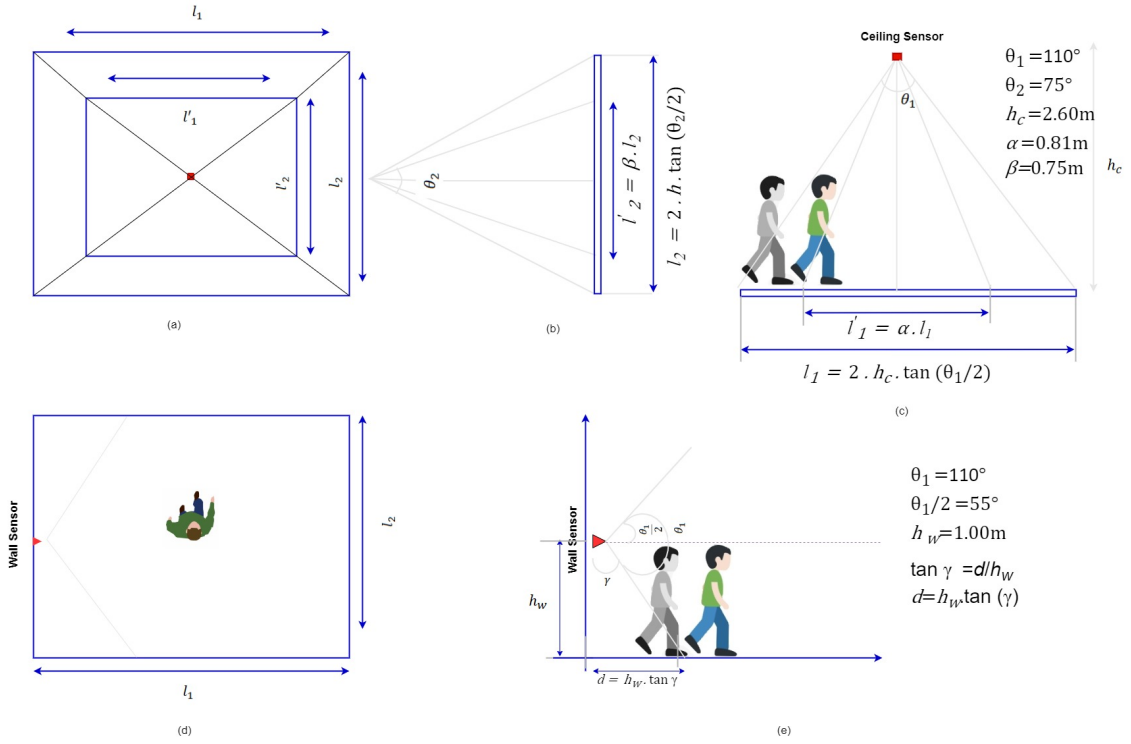


Figure 3-5: The area covered by the sensor and its detailed dimensions: (a) Top View of the ceiling sensor; (b) Side view of the ceiling sensor; (c) Front view of the ceiling sensor and its calculated dimensions; (d) Top view of the wall sensor; (e) Side view of the wall sensor and its calculated dimensions.

$$d = h_{\mathrm{w}} \tan \gamma \tag{3.4}$$

We experimented in 2 different rooms. The first room is a small closed space, with a little amount of sunlight entering from its single window. The Air Conditioner (AC) temperature in the room is set to 24° C. The second room is wider, has a large window allowing more sunlight to enter the room, and has an AC whose temperature is set to 22° C. Five different people, males and females of different ages, participated in the experiments. In each experiment, a single person is asked to perform various activities contentiously for 5 minutes. Data are collected by the sensors, which we use later on for classification. We conducted several experiments and collected enough data for training and evaluating the proposed approach.

Table 3.2: The frame counts for each activity in the training and the test data sets.

No.	Activity	Train data frames	Test data frames
0	Walking	1282	742
1	Standing	1174	956
2	Sitting	842	726
3	Lying	568	102
4	Action change	371	234
5	Falling	182	156

3.3.3 Data Collection

The two sensors kits run the same OS and script to collect the data. However, they collect the data independently from each other. This means that, even though they start simultaneously, a small time difference might occur. In such a case, we synchronize the data later on and discard accordingly a few frames from whichever sensor started before the other. Five people participated in our experiments, each performing different activities for over 5 minutes. Each 5-minute experiment generated over 2000 frames (per sensor), and therefore we collected in total more than $10,000$ frames. Each 5-minute experiment is referred to as a scenario. The collected 5 scenarios are split into a training data set and a validation data set. The training data set is obviously used to train our DL model, whereas the validation data set is used to evaluate the model. Three scenarios are used for training and two scenarios are used for validation.

As stated above, we collected over $10,000$ frames. Frames corresponding to the fractions of time where data are captured by one sensor and not the other, as well as frames where a person is located at the very edge of the coverage area are removed. Table 4.1 shows the distribution of the remaining frames per activity in both the training and validation sets.

Data collected using the IR array sensor differ drastically from that collected using typical RGB cameras. When we collect data using the IR array sensor, even for the same activity, the patterns of temperature distribution within the coverage area varies

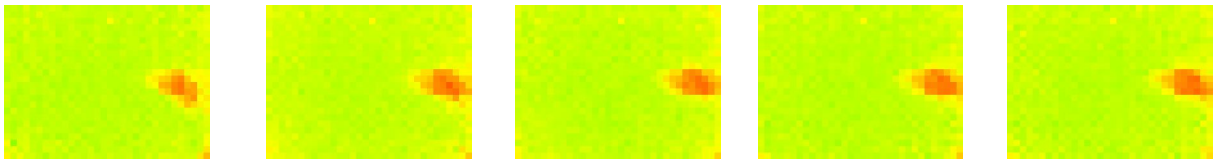


Figure 3-6: The temperature distribution of continuous frames of same activity at same position.

widely for different reasons. As can be seen, multiple factors lead to this difference in patterns. These factors can be summarized as follows:

- Blurriness: If we consider continuous frames of the same activity at a stationary position, the temperature distribution varies from one frame to another. This is shown in Fig. 3-6, and is due to the sensor characteristics the imprecision of the sensor capturing the temperature.

- Frames captured for the same activities, when the person is located at different positions, the shape of the pixels with high temperature differ, because the sensor captures different shapes.

- Attenuation of the temperature with distance: when the person moves to the edge of sensor's field of view, the temperature distribution attenuates since a lower temperature is captured by the sensor.

We address these issues in the current work as they affect the activity classification the most. Namely, the blurriness in the images conventionally lead to a drop in classification accuracy when using conventional methods, or when using a typical image classification CNNs.

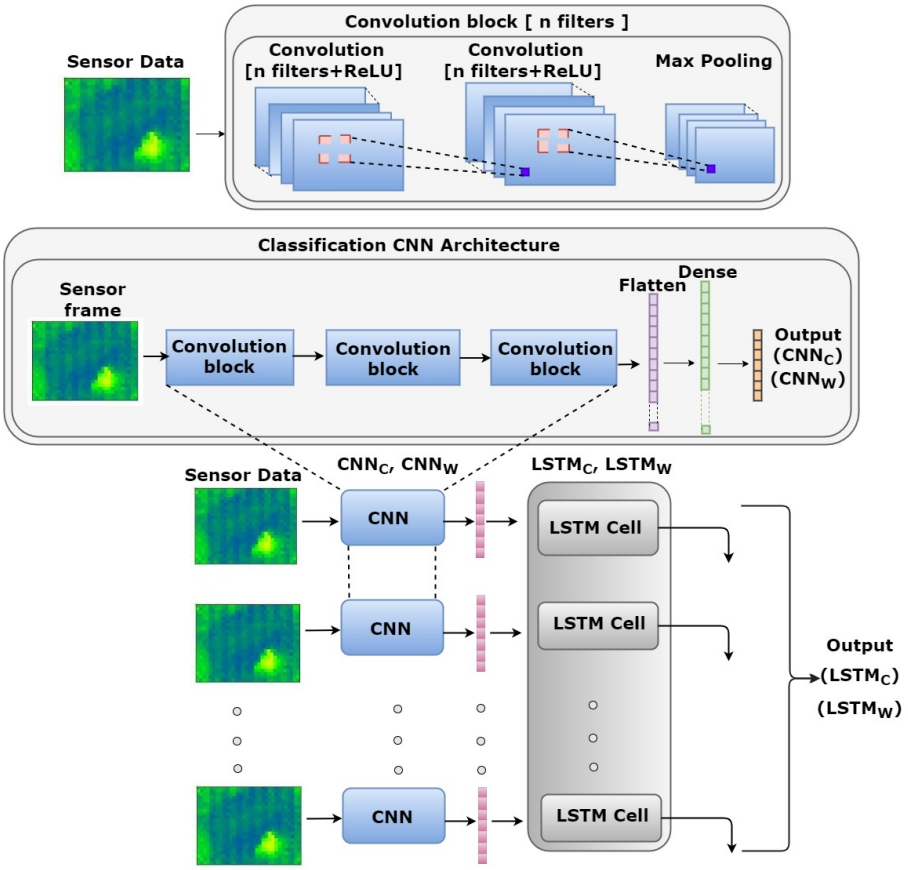


Figure 3-7: The General architecture of the neural network used for classification of both ceiling sensor data and wall sensor data.

3.4 System Architecture and Description

3.4.1 CNN and LSTM Architecture for Sensor Data Classification.

In the first step, we use data collected by each sensor individually to perform the activity detection. Frames collected by the sensor attached to the ceiling are classified using a CNN. Afterwards, the output of the CNN is passed to an LSTM which classifies a sequence of frames for more accurate judgment. In the same way, wall sensor data are classified using CNN and LSTM. The general architecture of the CNN and the LSTM is shown in Fig. 4-3. Both the ceiling sensor data and the wall sensor data are classified using this architecture.

Neural network As previously stated, throughout this work, we use a hybrid DL model to classify the different activities. Both the CNN and the LSTM networks are composed of the following typical layers:

- Convolution layer: a convolution layer typically takes as input either the raw data or the output of another layer, and applies a set of filters to output more "meaningful" data.
- Max pooling layer: this layer is usually used to reduce the dimensionality of the features extracted at a given previous layer by picking, for a subset of features, the one with the highest value.
- Flatten layer: this layer is used to flatten the data. In other words, it transforms a multi-dimensional matrix into a single vector.
- Dense layer: it aggregates all the features from the previous layers and maps them to the final features.
- LSTM cell: used for sequential classification of continuous input.

Convolution Layer The convolution layers consist of a set of filters with an activation function. The main function of a convolution layer is to get the input data and apply filterers to extract the features. In this CNN architecture, we used 2D-convolution layers. In this network, we use the term "convolution block" to refer to 2 consecutive 2D-convolution layers with Rectified Linear Unit (ReLU) activation function and filter size 3×3, followed by a MaxPooling layer. In our CNN, we have a total of six 2D-convolution layers, where every 2 consecutive layers are followed by a max pooling layer.

MaxPooling Layer The Maxpooling layer function is similar to that of the convolution layer as it also contains filters. It performs a specific function called pooling. MaxPooling is simply taking the maximum value of a subset of values from its input. This operation typicality reduces the dimensionality of the features. In our neural network, we used 2D-max pooling layer with a filter size 2×2.

Flatten Layer The flatten layer flattens the previous 2D layers output (which in return is a 2D matrix as well) by converting it into a single vector. This layer has no goal but to connect the 2D output to the fully-connected dense layer that comes after.

Dense Layer Dense layers are also referred to in the literature as "fully-connected layers". A dense layer aggregates all the information from the previous layer and maps them into a single feature vector used to identify the activity. The final dense layer outputs the class probability for the different activities. In other words, given an input frame, this last layer outputs a vector whose size is equal to the number of activities, where each value corresponds to the probability of that activity being shown in the frame.

Long Short Term Memory(LSTM) The LSTM is used for sequence classification of input data. It consists of there gates: the input gate, the forget gate, and the output gate. LSTM networks can retain information, allowing them to build a more accurate representation of the current state as a function of the previous ones, even ones far away in the past.

Activation Functions and Hyperparameters In this activity detection system, 2D-Convolution layers use ReLU activation function. This activation function does not activate all the neurons at the same time. Since the output of some neurons is set to zero, only a few neurons are activated making the network sparse, efficient, and easy for computation. The output dense layer uses a softmax function. We use a Stochastic Gradient Decent (SGD) optimizer to optimize the neural network. It reduces the chances of over fitting problem and is less computation-wise costly. For each model, we set dropout regularization between the layers with a probability equal to 0.2. Batch normalization is used to accelerate the training process. These are the details of the hyper parameters used in all the models of our activity detection system.

Our neural networks are designed based on Convolution-LSTM [186] and Siamese Neural Network architecture [187]. This is a common family of neural networks for

sequential activity classification. However, the architecture that we propose, as it stands is novel and has been designed taking in mind 3 factors: 1) the type of input data (i.e., sequences of 32×24 images) which are very low resolution, 2) the requirement in terms of performance: more complex neural networks might increase the accuracy slightly but not much, and less complex ones have a remarkable performance degradation, and 3) the complexity itself: we expect our model to run on low computation devices such as the Raspberry Pi (which we used to collect the data). A more complex neural network architecture might end up being very costly for a negligible performance improvement.

3.4.2 CNN and LSTM Architecture for Combined Sensor Data Classification.

The architecture of the CNN used for the classification of the combined data (i.e., data collected from the ceiling sensor and data collected from the wall one) is shown in Fig. 3-8. The parameters of the different layers of the neural networks (both the CNN and the LSTM) are the same as explained in the previous subsection. The outputs of the first dense layers of the two sub-networks are concatenated and are connected to a single dense layer whose size is equal to the number of activities. This dense layer obviously outputs the probabilities of the activities.

The combined CNN output is passed to the LSTM whose detailed architecture is shown in Fig. 3-9. The input to the LSTM is a vector in the time domain whose size is equal to 5. Each time step consists of a vector whose size is equal to 6, which is the output of the CNN.

The combined CNN+LSTM and combined CNN neural network are designed based on convolution-LSTM and Siamese neural network, respectively. Our CNN neural network architecture automatically learns the features and the weight of individual frames. The layers of the CNN are then frozen, and the LSTM is trained to use the output of the CNN to run the classification. This has led to a good prediction result compared to the CNN when used alone. These kinds of architecture have sev-

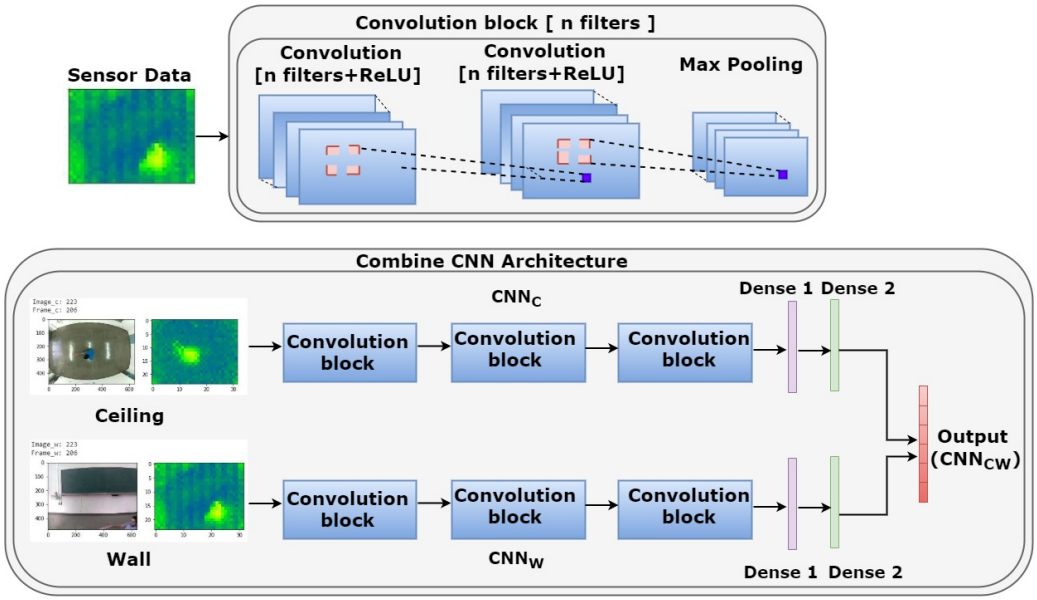


Figure 3-8: The architecture of the combined CNN for classification.

eral advantages: On the one hand, the CNN is more robust in classifying imbalanced data, and able to achieve high accuracy of classification on its own. On the other hand, some activities require observation over an extended period of time to detect the motion. Here the LSTM has a higher potential in detecting such activities.

3.5 Experimental Results

3.5.1 Performance Evaluation Metrics

We use precision, recall, F1-score, and accuracy as metrics for evaluating the efficiency of the proposed activity detection approach. The True Positive (TP), False Positive (FP), True Negative (TN), and, False Negative (FN) values are reported in the confusion matrix. The evaluation metrics are based on the following formulas:

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{3.5}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{3.6}$$

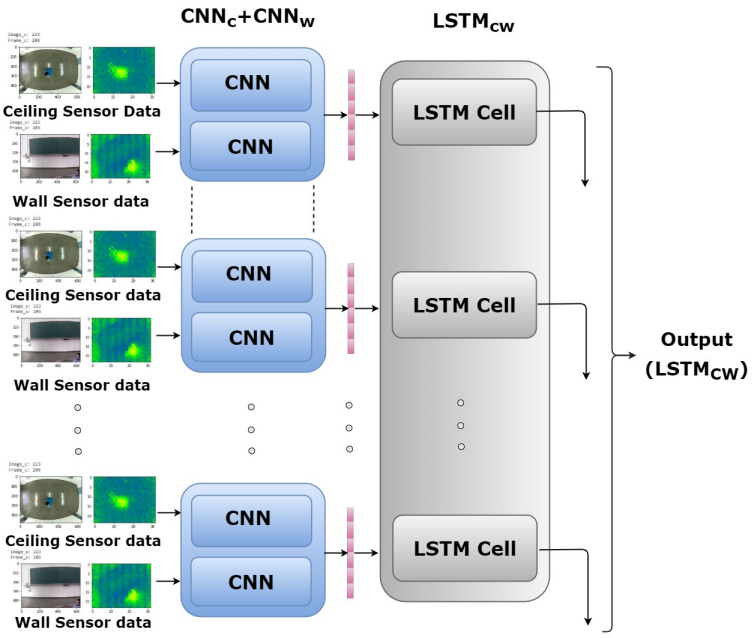


Figure 3-9: The architecture of the combined CNN and LSTM for classification.

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{3.7}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN}. \tag{3.8}$$

We obtained good results from each of the models. However, it is essential to show the ability to detect the activities and robustness against false positives. For this, we use precision and recall. Precision measures the correctly classified instances of a given class relative to all the instances classified as belonging to that class. Recall measures the number of correctly classified instances of a given activity relative to all its instances. F1-score is the harmonic mean of both precision and recall.

3.5.2 CNN Classification Results

The confusion matrix of the classification of the ceiling sensor data using CNN is shown in Table 3.3. Based on the observation of this confusion matrix, sitting and standing activities are the most confused ones; and walking and action change activ-

Table 3.3: The confusion matrix of the classification of the ceiling sensor data.

Class	**Classified as**					
	0	1	2	3	4	5
Walking-0	**717**	9	0	0	14	2
Standing-1	6	**913**	26	0	11	0
Sitting-2	11	23	**678**	0	14	0
Lying-3	0	7	12	**83**	0	4
Action change-4	19	0	1	6	**208**	0
Falling-3	4	0	3	0	0	**149**

Table 3.4: The precision, recall and F1-score for classification of ceiling sensor data using CNN for each activity.

Activity	**Precision**	**Recall**	**F1-Measure**
Walking	0.95	0.97	0.94
Standing	0.96	0.96	0.96
Sitting	0.94	0.93	0.93
Lying	0.93	0.78	0.85
Action change	0.89	0.94	0.91
Falling	0.96	0.96	0.96

ities confusion comes second. From this, we conclude that there is confusion between the sitting and standing activities when classifying ceiling sensor data using CNN.

Next, the performance evaluation for classification of the ceiling sensor data using CNN is shown in Table 3.4. Walking and action change are misclassified activities, as we can see from our previous observations from the confusion matrix. For instance, despite its high recall, walking activity has low precision. This leads us to believe that the CNN's performance for the walking activity needs to be improved. Falling and sitting activities have the highest classification performance, with falling reaching the highest precision and F1.

The results of the classification of the wall sensor data using CNN is shown in Table 3.5. The confusion matrix shown here, illustrates that this model does not perform well for many activities. Misclassification of the sitting, standing, and walking activities is very high owing to the limitations in the detection accuracy arising from the activity being carried out of the wall sensor's coverage range. These limitations

Table 3.5: The confusion matrix of the classification of the wall sensor data using CNN.

Class	**Classified as**					
	0	1	2	3	4	5
Walking-0	**719**	0	0	5	11	7
Standing-1	0	**917**	19	17	3	0
Sitting-2	0	17	**674**	23	12	0
Lying-3	0	12	27	**63**	0	0
Action change-4	12	0	0	1	**217**	4
Falling-5	6	0	2	0	3	**145**

Table 3.6: The precision, recall and F1-score for classification of wall sensor data using CNN for each activity.

Activity	**Precision**	**Recall**	**F1-Measure**
Walking	0.98	0.97	0.97
Standing	0.97	0.96	0.96
Sitting	0.93	0.93	0.93
Lying	0.58	0.62	0.60
Action change	0.93	0.89	0.91
Falling	0.93	0.93	0.93

need to be overcome for the model to perform well. In addition to the confusion matrix, the detailed performance evaluation of this model is shown in Table 3.6.

Based on the results so far, it is clear that the detection of some activities such as sitting and falling is good when using the ceiling sensor, whereas the detection of some other activities such as action change is better when using the wall sensor. Clearly, there is a need for improvement of the detection of all the activities in a collective manner. To do so, we combine both the data collected by the ceiling sensor and those collected by the wall sensor and perform the classification on these combined data using CNN.

The results of the classification of the combined data using CNN is presented in Table 3.7. From these results, we clearly see that the overall misclassification of all the activities has been reduced as compared to the previous results when using individual sensor data for classification using CNN. Walking, standing, sitting, falling,

Table 3.7: The confusion matrix of the classification of the combined sensor(s) data using CNN.

Class	Classified as					
	0	1	2	3	4	5
Walking-0	**721**	0	0	13	0	8
Standing-1	0	**940**	9	7	0	0
Sitting-2	0	12	**705**	9	0	0
Lying-3	7	4	13	**73**	5	0
Action change-4	5	0	0	7	**220**	2
Falling-5	3	0	0	0	2	**151**

Table 3.8: The precision, recall and F1-score for classification of the combined sensor data using CNN.

Activity	Precision	Recall	F1-Measure
Walking	0.98	0.97	0.97
Standing	0.98	0.98	0.98
Sitting	0.97	0.97	0.97
Lying	0.67	0.72	0.69
Action change	0.94	0.89	0.91
Falling	0.94	0.97	0.95

and lying activities have good detection measures compared to the individual sensor data results. The performance evaluation of this classification is shown in Table 3.8. Walking, standing, and sitting activities show a remarkable improvement in terms of detection based on the precision and recall values.

3.5.3 CNN and LSTM Classification Results

In this section, we discuss the results of the hybrid DL model. In this model, the output of the CNN is passed to the LSTM for sequence classification.

The confusion matrix of the sequential classification of the ceiling sensor data is shown in Table 3.9. From this confusion matrix, it can be inferred that standing and lying are the most misclassified activities. The performance evaluation metrics for this experiment are shown in Table 3.10. Based on the results of this experiment's confusion matrix and previous results regarding the classification of ceiling sensor

Table 3.9: The confusion matrix of the classification of the ceiling sensor data using CNN and LSTM.

Class	**Classified as**					
	0	1	2	3	4	5
Walking-0	**721**	7	0	0	11	3
Standing-1	0	**920**	14	22	0	0
Sitting-2	8	7	**711**	0	0	0
Lying-3	0	5	14	**79**	0	4
Action change-4	6	0	0	0	**221**	7
Falling-3	1	0	0	0	3	**152**

Table 3.10: The precision, recall and F1-score for classification of ceiling sensor data using CNN and LSTM.

Activity	**Precision**	**Recall**	**F1-Measure**
Walking	0.98	0.97	0.97
Standing	0.96	0.96	0.96
Sitting	0.96	0.98	0.97
Lying	0.78	0.77	0.77
Action change	0.94	0.94	0.94
Falling	0.92	0.97	0.94

data using CNN, it can be concluded that lying is the only activity that requires an improvement in detection.

The confusion matrix of the classification of the wall sensor data using CNN and LSTM is shown in Table 3.11. This result in relation to the results obtained from the classification of the wall sensor data using CNN can be used to reduce the misclassification rate of lying and sitting activities with the improvement in detection for the other activities in case of the current results. Furthermore, Table 3.12 shows the performance evaluation metrics of the classification using this model. It can be observed that the detection of walking and standing activities performed well in this model. However, detection of lying and sitting activities is still low due to the same reason(s) described in the classification of the wall sensor data using CNN, i.e., the limitation in activity detection due to the subject being out of the sensor's peripheral vision.

Table 3.11: The confusion matrix of the classification of the wall sensor data using CNN and LSTM.

Class	**Classified as**					
	0	1	2	3	4	5
Walking-0	**724**	11	0	0	7	0
Standing-1	0	**923**	24	9	0	0
Sitting-2	7	13	**706**	0	0	0
Lying-3	0	4	26	**72**	0	0
Action change-4	11	0	0	0	**219**	4
Falling-5	5	0	0	0	11	**206**

Table 3.12: The precision, recall and F1-score for classification of wall sensor data using CNN and LSTM.

Activity	**Precision**	**Recall**	**F1-Measure**
Walking	0.97	0.98	0.97
Standing	0.97	0.92	0.94
Sitting	0.93	0.97	0.95
Lying	0.89	0.71	0.79
Action change	0.94	0.89	0.91
Falling	0.97	0.90	0.93

Based on the above results obtained so far for sequential classification (CNN+LSTM), certain activities are better detected when using the ceiling sensor data, whereas others are better detected when using the wall sensor data. Therefore, there is still room for improvement in the activity detection when using the hybrid model, as our ultimate goal is to have a single model performing well for all the activities. This is best illustrated in the case of detection of lying and sitting activities which is low when the wall sensor data are used compared to when the ceiling sensor is used.

The confusion matrix of the sequential classifier (CNN+LSTM) is presented in Table 3.13. The results reported in the confusion matrix show an improvement in the detection of all the activities, compared to the case where we used only the CNN. However, lying activity is still relatively poorly detected, requiring further improvement. That being the case, to improve the detection performance of all the activities but particularly that of lying, we combine both the ceiling sensor data and

Table 3.13: The confusion matrix of the classification of the combined sensor(s) data using CNN and LSTM.

Class	Classified as					
	0	1	2	3	4	5
Walking-0	**727**	8	0	0	5	2
Standing-1	0	**939**	8	9	0	0
Sitting-2	3	9	**714**	0	0	0
Lying-3	0	0	19	**73**	11	0
Action change-4	13	0	0	0	**219**	2
Falling-5	1	0	0	0	2	**153**

Table 3.14: The precision, recall and F1-score for classification of combined sensor(s) data using CNN and LSTM.

Activity	Precision	Recall	F1-Measure
Walking	0.98	0.98	0.98
Standing	0.98	0.98	0.98
Sitting	0.96	0.98	0.97
Lying	0.89	0.71	0.79
Action change	0.94	0.89	0.87
Falling	0.97	0.98	0.97

wall sensor data and perform the classification using CNN and LSTM.

The performance evaluation of the classification of the combined sensor data using CNN and LSTM is shown in Table 3.14. The results indicate that this model performed the best as compared to the other models that have been discussed so far. In particularly noticeable in the case of walking, standing, falling, and action change activities.

3.5.4 Overall Performance

Accuracy is defined as the correctly classified instances over all the instances of all the activities. We downscale our dataset to 8×8 and run it on the previous existing conventional machine learning models. We classified different activities using various models that have been used in the conventional works. Table 3.15 shows the comparison of classification accuracy for these models. Based on this table, we observe that

Table 3.15: Comparison of the classification accuracy of our models with those in the conventional work.

Methods	**No. of the sensor**	**Position and classification accuracy**		
		Ceiling	Wall	Combine ceiling and wall
SVM [15]	1	0.72	-	-
k-NN [55]	1	0.84	-	-
SVM [67]	2	✓	✓	0.90
CNN	2	0.94	0.93	0.96
CNN+LSTM	2	0.96	0.95	0.97

the combined sensor data classification using CNN and LSTM model performed the best and reached over 0.97 accuracy.

Chapter 4

AD Systems Using Single IR Array Sensor

4.1 Introduction

AD systems using a wide-angle IR array sensor with advanced DL based CV techniques. The wide angle IR sensor used produces noisy and blurry images, making it difficult to accurately identify the activity. Due to environmental factors such as sunlight, temperature, heat objects, etc., the amount of noise and blurriness in the IR images is high. This makes it difficult to correctly identify the activity. Making a generalized model that works for different environments and that is robust to differences between the conditions where/when it was trained and those where/when it is deployed is also a very challenging task. To address all these challenges and to build a robust activity detection system, we employ more advanced DL techniques. In this approach, we use a single IR sensor placed on the ceiling and collect data with various resolutions (i.e., 24×32, 12×16, and 6×8). To faithfully increase the resolution and enhance the low quality of the collected data, we use two techniques referred to as SR [188] and image denoising [189]. By enhancing the quality of the collected images, not only do we improve the activity detection accuracy, but we also make it more robust to changes in the environment, namely ones related to the temperature and the presence of noise sources. We use two mainstream types of DL classifiers, namely

a CNN and a combination of a CNN and a LSTM. The classification process goes as follows. First, all the individual images are classified using CNN. The CNN learns the appropriate weight in the convolution part of the network and performs a rough classification of activities. Second, the CNN output is passed to the LSTM, which performs a more robust classification by taking into account the temporal component. Nonetheless, since it is difficult to collect the data in many environments, we use a technique referred to as data augmentation [190] to generate artificial data that mimics real ones. For this sake, we employ a particular type of neural network conceived for this task known as Conditional Generative Adversarial Networks (CGAN). The use of the aforementioned DL techniques leads to a noticeable improvement in the AD.

4.2 Framework of AD Systems Using Advance DL Based CV Techniques

A flowchart of the overall framework is shown in Fig. 4-1. The data collected by the sensor have 24×32, 12×16, and 6×8 resolution. We apply the SR and denoising techniques to the LR data. The HR data 24×32 used to generated synthetic data using CGAN to diversify the samples and cover potentially important missing samples. The synthetic data are used to train the CNN model. We classify the individual frames using a CNN. The output of the CNN is passed to the LSTM with a time window size of five frames to improve the classification accuracy. Finally, we compare all the data classification performance.

4.3 Data Collection

Experiments are conducted by placing IR sensor on the ceiling subjects were ask to perform the six different activities (i.e., walking, falling, standing, sitting, lying, and the action change is define as transition between the activities.) under the sensor coverage area. We conducted the experiment in three different rooms with following

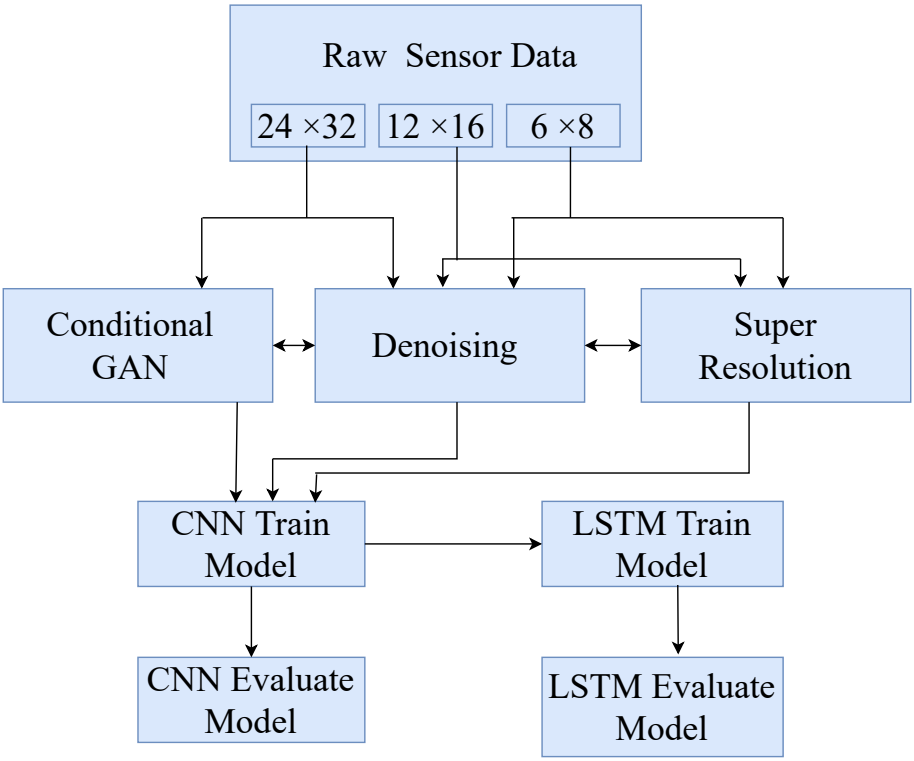


Figure 4-1: A flowchart of the proposed system.

different environment

- The first room is a small, closed space with only one window that lets in little light. The temperature in the room has been set to 24°C.
- The second room is larger, brighter, and equipped with an air conditioner whose temperature is set to 22°C.
- In comparison to the other rooms, the third room is a little dark and its air conditioner temperature is set to 24°C.

Some examples of the data collected in different environments with different resolutions are shown in Fig. 4-2

In total we collected 12 scenarios of data. Each scenario lasts for five minutes. The camera and the sensor both collected data at 8 frames per second. One scenario is defined as 5 minutes of continuous activities. Each scenario includes all the activities (i.e., walking, falling, standing, sitting, lying, and the action change referred

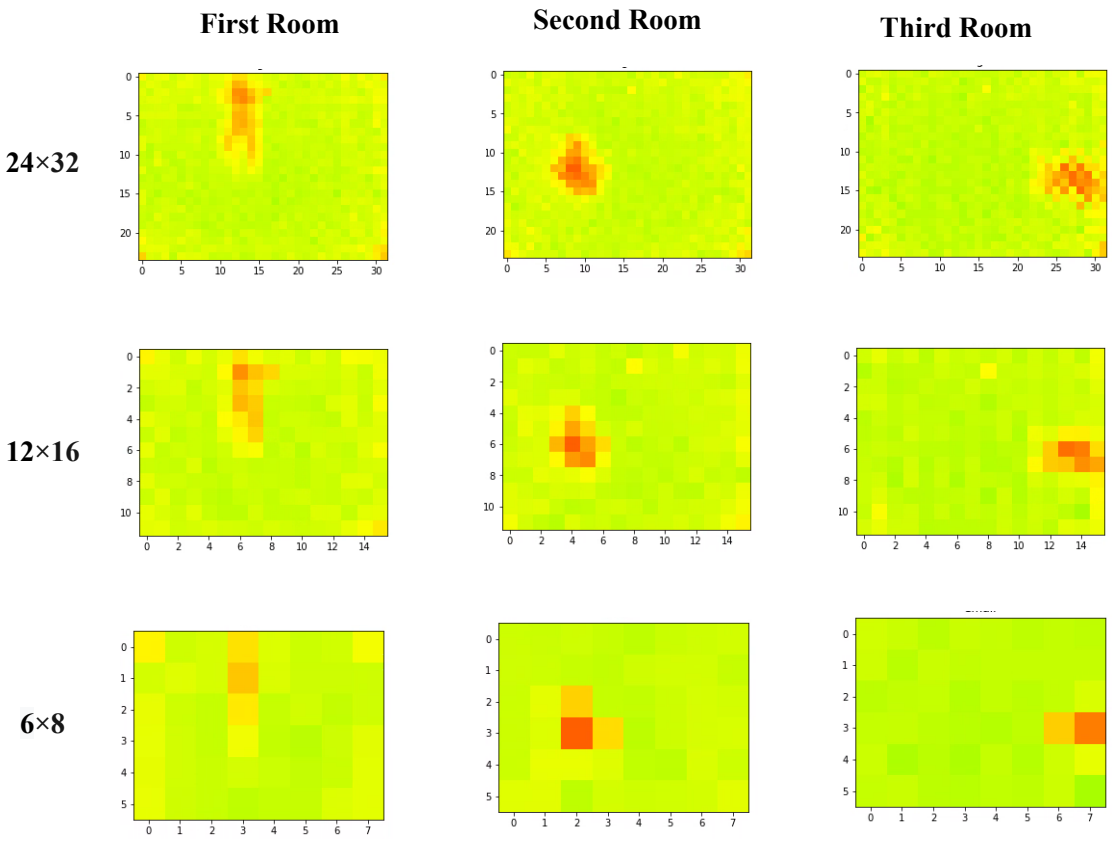


Figure 4-2: Some examples of the raw data collected in different environments with different resolution.

transition between the activities.). Out of the 12 scenarios, we used 8 for training and the remaining 4 to test the model of our proposed approach. Table 4.1 shows the distribution of frames showing the different activities in the training and test data sets. Raw data from the sensor is 24×32 resolution to obtain the LR data we downscale the images to 12×16 and 6×8 resolutions.

4.4 DL based CV Techniques

4.4.1 Super-Resolution

The SR technique is used on LR data (6×8, 12×16), to learn how to upscale them back to the HR resolution 24×32. By doing so, we can use low-end cheaper sensors

Table 4.1: The frame counts for each activity in the training and testing data sets.

No.	Activity	Training data frames	Testing data frames
0	Walking	5456	2351
1	Standing	1959	882
2	Sitting	3102	1566
3	Lying	2486	647
4	Action change	1961	939
5	Falling	613	264

that collect the data naively at these low resolution (i.e., 6×8 and 12×16) and apply the trained SR model to upscale them faithfully to a higher resolution, then perform the classification. By doing so, it is possible to improve the classification accuracy of frames collected by low-end sensor to match (or get as close as possible to) the HR 24×32 pixel frames collected by higher-end more expensive ones.

In our work, we use the Fast Super-Resolution convolution Neural Network (FSRCNN) [188] to improve image quality. The architecture of the neural network is depicted in Fig. 4-3. It is based on a shallow network design that reproduces images faster and more clearly. The FSRCNN is made up of five components:

- Feature extraction
- Shrinking
- Non-linear mapping
- Expanding
- Deconvolution

Feature Extraction

The low-resolution image's overlapping patches are extracted and represented as a high-dimensional feature vector. It is accomplished through the use of n_1 convolution filters with kernel sizes equal to 5×5.

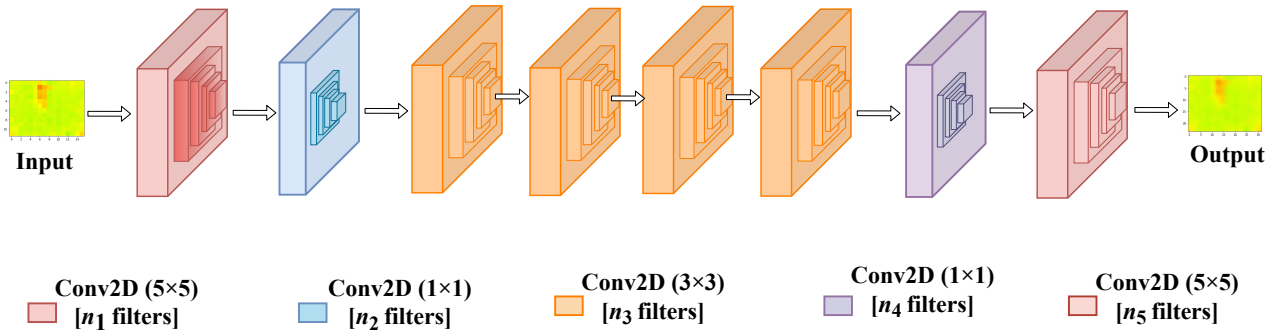


Figure 4-3: The architecture of the neural network used for Super-Resolution.

Shrinking

To reduce the feature dimension, a shrinking layer is added after the feature extraction layer. This will help reduce the computational complexity. In this convolution layer, a set of n_2 filters of size 1×1 is used to linearly combine the low-resolution features.

Non-Linear Mapping

Non-linear mapping is one of the vital part of the SR process. The purpose of the non-linear mapping is to map the feature vector to a higher dimensional space. This higher dimensional space contains richer information that could be mapped to the expected output vector. In other words, it is responsible for generating enough context to reconstruct the high-resolution image. The number of sub-blocks in the non-linear mapping block and the filter size used highly influence the neural network's performance. In our current work, we adopted a convolution layer with n_3 filters of size 3×3. The total number of sub-blocks is represented as $m = 3$.

Expanding

The shrinking layer reduces the dimension of the low-resolution feature. The quality will be poor if we generate HR image directly from the low-resolution feature dimension. This is why we add the expanding layer after the mapping block. We use a convolution layer with n_4 filters of size 1×1 to maintain consistency with the shrinking layer

Deconvolution

A deconvolution layer is used to upscale and aggregate the previous layer's output, resulting in a more accurate representation of the data. Unlike the convolution layer, the deconvolution layer expands the low-resolution into higher dimension data. More precisely, this is determined by the stride size, as a stride of size 1 with padding would yield information of the same size, whereas a stride of size k will yield condensed information of size $1/k$. Deconvolution with stride expands the input data so that the output image can be 24×32 resolution.

Activation Functions and Hyperparameters

In FSRCNN, a new activation function introduced called Parametric Rectified Linear Unit (PReLU) for better learning. The activation threshold of PReLU is different from that of conventional ReLU. PReLU's threshold is learned through training, whereas ReLU uses a fixed 0 as a threshold, mapping all negative values to zero. This is essential for both training and later estimating the architecture's complexity.

We use our neural network's total number of parameters as an indicator to estimate its complexity. To recall, our network is basically composed of a set of convolutions followed by a single deconvolution. In addition to that, we include the number of PReLU parameters.

To measure the total number of parameters of the neural network, we use the following equations which measure the total number of parameters in a convolution layer (C_{sr}), and the parameter in the PReLU layer (A_{sr}):

$$C_{sr} = ((mnp) + 1)k, \tag{4.1}$$

$$A_{sr} = hwk, \tag{4.2}$$

where m and n are the width and height of each filter, respectively, p is the number of channels, k is the number of filters used in the layer, and h and w are the input image's height and width, respectively.

As a result, the total number of parameters in the network is 21,745 for the input images are 8×6 and 47,089 for the input images are 16×12.

An example of a 24×32 image, its low resolution version to 12×16 (resp. 6×8) and the reconstructed SR one are given in Fig. 4-4.

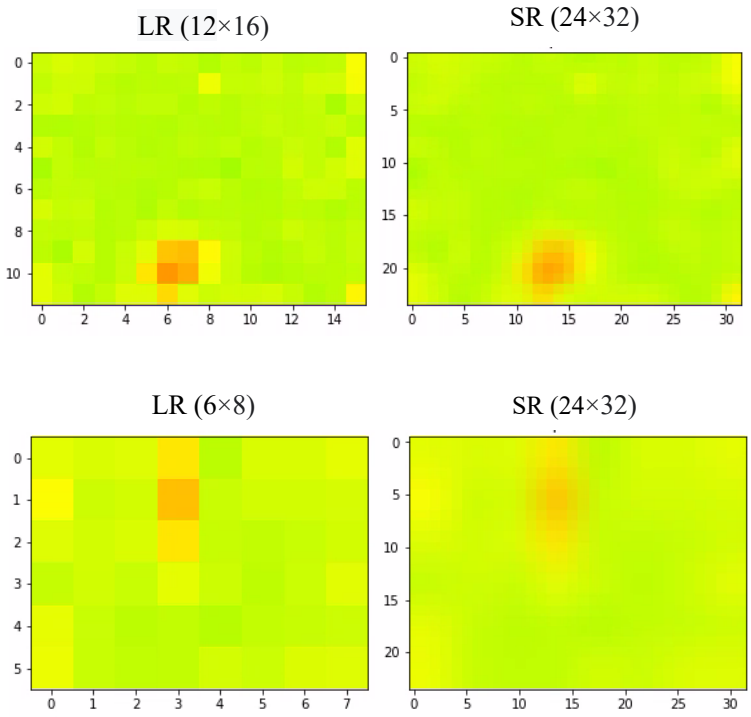


Figure 4-4: The output of SR technique applied to 12×16 frame and 6×8 frame.

4.4.2 Denoising

Denoising refers to the process of restoring an image that has been contaminated by additive noise. Due to their ability to learn very fine patterns in an image, deep convolution networks have proven to be highly effective in denoising images in recent years. One of the image restoration techniques is the Deep Image Prior (DIP) [189]. This technique demonstrates that the network structure is adequate to restore the original image from the degraded image. Pretrained networks or large image datasets are not required for this technique. It operates directly on the degraded images and learns internally what makes noise and what makes useful pixels.

Generally speaking, the most commonly used methods for image restoration in CV are learned prior [191] and explicit prior [192]. Learned-prior is a simple method for training a deep convolution network to learn how to denoise images by training on a data set. It takes noisy images as training data and clean images as ground truth, and trains the network to reconstruct the clean image from the noisy one. In the explicit-prior method noises are mathematically calculated and removed. DIP bridges the gap between these two popular methods by constructing a new explicit prior using a convolution neural network.

The DIP structure is based on the U-Net [193] type neural network shown in Fig. 4-5 with multiple downstream and upstream steps and skip connections, each of which consists of a batch normalization and an activation layer. Random noise is fed into the network. The target is the image that has been tainted by the use of a mask. The loss is calculated by applying the same mask to the output image x^* and comparing it to the noisy image. This implies that the loss function does not explicitly drive the noise/corruption repair (as it is re-applied before computing the loss). This is due to the neural network's implicit behavior. When the network attempts to optimize toward the corrupted image. The neural network contains the parameterized weight θ. Based on the θ, the neural network will use gradient descent optimization to find the optimized weight θ^{k+1}.

$$\theta = \underset{\theta}{\operatorname{argmin}} E(f_\theta(z); x_0), \tag{4.3}$$

$$\theta^{k+1} = \theta^k - \alpha \frac{\delta E(f_\theta(z); x_0)}{\delta \theta}, \tag{4.4}$$

$$x^* = f_\theta(z), \tag{4.5}$$

where x_0 is noisy image and z is random noise. Here $E(f_\theta$ (z); $x_0)$ is a data term usually used in the denoising problem. The f_θ is convolution neural network based encoder-decorder parameterized by the weight θ. The result of denoised images is shown in Fig. 4-6. In our work, we applied the denoising DIP technique on 24×32,

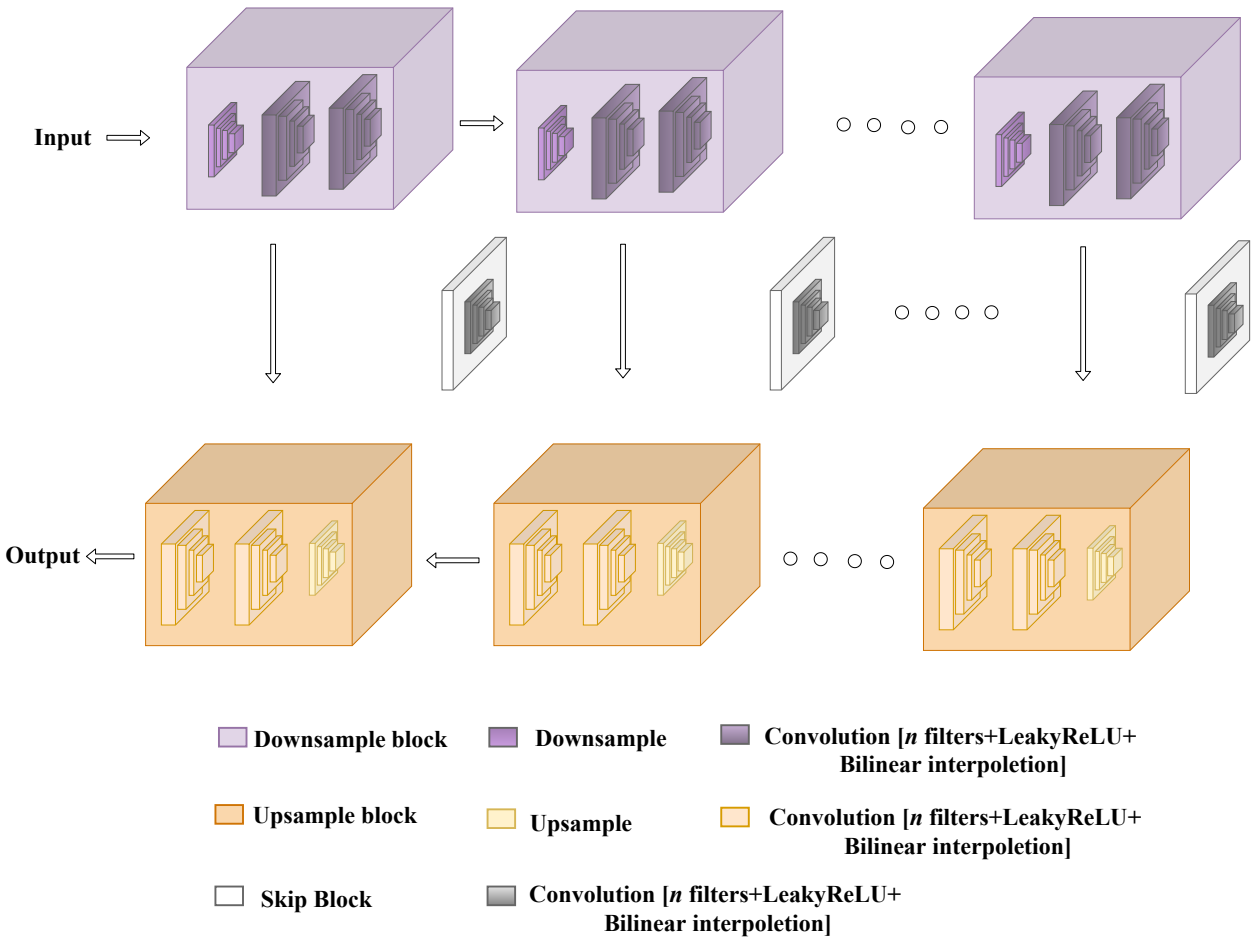


Figure 4-5: The architecture of the neural network used for denoising.

12×16, and 6×8 data.

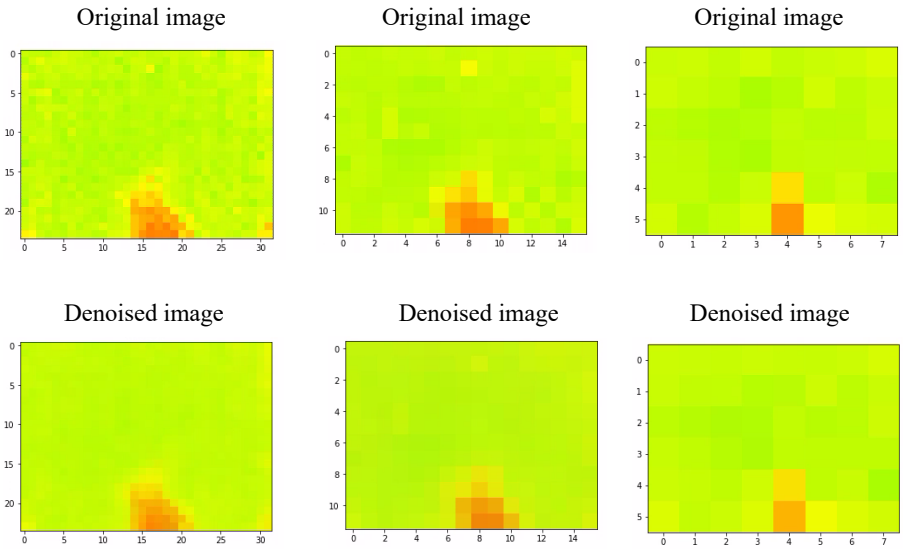


Figure 4-6: The outputs of denoising technique applied to a 24×32, 12×16, and 6×8 frame.

4.4.3 Conditional Generative Adversarial Networks (CGAN)

When training neural networks, a common technique referred to as "data augmentation" is used to address some of the issues related to the nature and amount of data used for training. Data augmentation refers to the process of generating artificial (or synthetic) data to enlarge the size of the training set. The synthetic data improve the classification result and strengthen the system's ability to work in various environments. The most advanced DL technique for data augmentation is the conditional generative adversarial neural network (CGAN) [190]. CGAN is a generative model for supervised learning. The labeled data are used to train and generate synthetic data based on the number of classes. The CGAN structure is comprised of two neural networks: a generator G and a discriminator D, as depicted in Fig. 4-7. x is the real image, $p_{data}(x)$ and $p_z(z)$ denote the distribution of the real and the synthetic samples, respectively. A random noise z is taken from prior distributions with the label y and is used as an input to the generator known as a

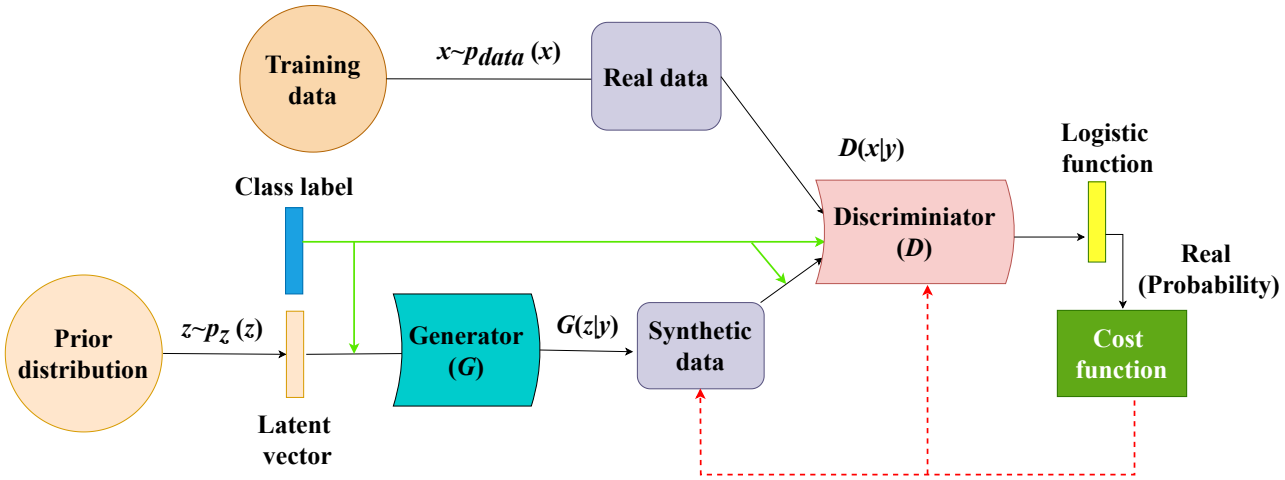


Figure 4-7: The architecture of data augmentation technique (CGAN).

latent vector ($(z \mid y) \sim p_z(z \mid y)$). The generator aims to create, out of the input noise, samples with more complex distribution $G(z \mid y)$ that similar to that of the real ones (i.e., x) for the given class y. $(x \mid y)$ and $(z \mid y)$ are the real image with label and random noise with label, respectively. In the meantime, the discriminator should distinguish between real samples ($(x \mid y) \sim p_{data}(x \mid y)$) and the generated samples($G(z \mid y) \sim p_z(z \mid y)$). Back-propagation optimizations are used to train both networks, and they are completely independent of one another. The optimization of the generator using the discriminator's predictions about the samples it generated. The discriminator is trained using the generator's synthetic data. This optimization uses CGANs' training cost function, min max loss, as shown in the equation below.

$$\min_G \max_D = E_{x \sim p_{data}(x)} \left[\log \left(D\left(x|y\right)\right)\right] + E_{z \sim p_z(z)} \left[\log \left(1 - D\left(z|y\right)\right)\right] \tag{4.6}$$

After several iterations of the two training techniques described above, the generator learns to generate more sophisticated samples that do resemble the real ones, and the discriminator learns how to identify the slight variation between the real and synthetic data. To reduce the cost function of each network and optimize its internal

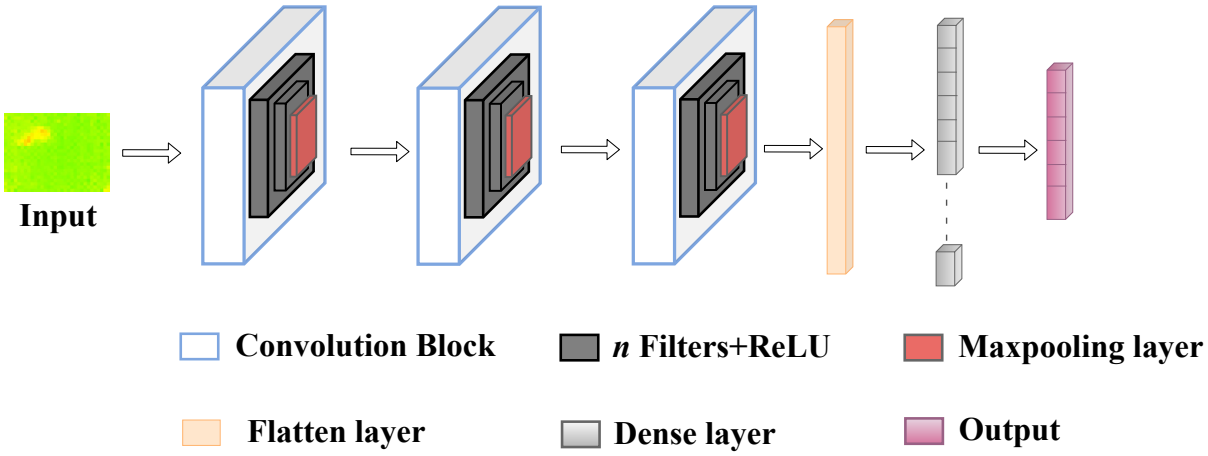


Figure 4-8: The architecture of the CNN used for classification.

weights, a gradient step with back-propagation is performed at each iteration.

4.5 Activity Classification

The classification neural network's architecture is illustrated in Fig. 4-8 and Fig. 4-9. The classification consists of two stages:

- In the first stage, the sensor's raw data are given as input to the CNN that classifies the individual frames and produces the first output.
- In the second stage, we perform the sequence classification using the LSTM. The output of the CNN is given as input to the LSTM with a window size equals to five frames. The LSTM produces the sequence classification output.

Our neural network architecture consists of six 2D-convolution layers and two fully connected layers. Each convolution layer uses filters with a kernel size equal to 3 and has a ReLU activation function. Every two 2D-convolution layers are followed by a 2D-Maxpooling layer whose kernel size is set to 2. The output of the sixth convolution layer is flattened, and is connected to a dense layer with a ReLU activation function. In the final dense layer, the activation function is sigmoid. The output of the CNN is given as input to the LSTM network.

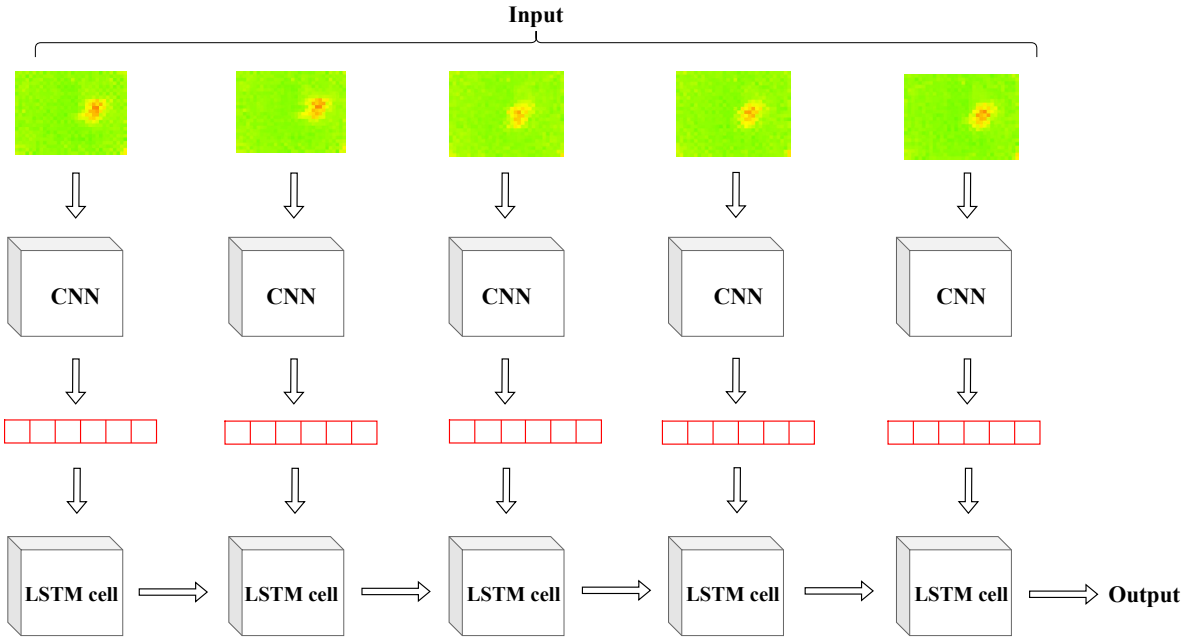


Figure 4-9: The architecture of the CNN+LSTM network used for classification.

To evaluate the complexity of the neural network, we measure the total number of parameters of the neural network. We use the following equations which measure the total number of parameters in a convolution layer (C_p), that in a dense layer (D_p), and that in an LSTM layer (L_p):

$$C_p = ((w_f h_f p) + 1)c, \tag{4.7}$$

$$D_p = ((sn) + 1), \tag{4.8}$$

$$L_p = 4((i_l + 1)d_l + d_l^2), \tag{4.9}$$

where w_f, h_f, and c represent the width, height, and the number of channels of each filter, respectively, f represents the number of filters in the convolution layer, s represents the size of the dense layer, and n represents the number of neurons in the previous layer. The i_l and d_l are the input and output sizes of the LSTM neural network, respectively. The total number of parameters is about 189K in the

Table 4.2: A comparison between the total number of parameters of the neural networks used in the current work and those of the state of the art neural networks used for image classification.

Model	Parameters
ResNET [194]	21 Million
VGG16 [195]	138 Million
CNN	189 Thousand
CNN+LSTM	568 Thousand

CNN, and about 568K in the LSTM network. Compared to the existing pre-trained models like ResNet [194] (21M parameters) and VGG16 [195] (138M parameters), our model is lightweight and can easily run on low-end computational devices such as the Raspberry Pi. The total number parameters of the neural networks proposed in this work is shown in Table 4.2 along with that of some of the state of the art neural network architectures. For its size and weight, the proposed architectures provide a very good classification performance.

4.5.1 Further Model Optimization Using Quantization

In the realm of DL, Quantization [196,197] refers to the concept of using low bit width (conventionally 8-bits) numbers to represent the weights within the neural network, rather than using floating numbers, which occupy much more space, and are more computation costly. These operations with low bit width numbers, such as integers, are the lighter from a computer's perspective.

With that in mind, to achieve a high accuracy for our models while keeping their computational demands as low as possible, we use this concept of quantization as introduced in [196, 198, 199] to reduce the size of our model. The purpose of weight quantization is to replace high weights with low weights without modifying the network's architecture. As a result, approximated weights are used for compression. There is a trade-off between weight quantization and classification accuracy because precise weight is given up for low memory space. Weight sharing typically utilizes the

same weight rather than retraining parameters. This significantly reduces computational costs. We use a quantization aware training [200, 201], which has a lower loss in quantization. However, it is important to emphasize, that despite its contribution to the minimization of the model size and the computation cost, the accuracy of the model when using quantization drops compared to that when using the original weights in the model after training. Quantization aware training (QAT) [201] works by applying a fake quantized 8-bit weight float to the input. The training is then operated normally as it deals with floating point numbers, even though it emulates operations with low bit width numbers. Once the training is complete, the information stored during the fake quantization are used to convert the floating-point model to an 8-bit quantized model.

4.6 Experimental Results

4.6.1 CV Performance Results

To evaluate the network model's performance in image SR and denoising, this paper utilizes a widely used image quality metric, namely the Peak Signal-to-Noise Ratio (PSNR). PSNR is commonly used to objectively evaluate image quality. It is defined as the ratio between the maximum power of the effective signal and the power of the noise in the signal. PSNR is measured in decibels (dB), and its mathematical expression is

$$\text{PSNR} = 10 \times \log_{10}\left(\frac{2^n - 1}{\text{MSE}}\right), \tag{4.10}$$

$$\text{MSE} = \frac{1}{mn}\sum_{i=0}^{n}\sum_{j=0}^{m}|X_{ij} - Y_{ij}|^2. \tag{4.11}$$

Here, MSE stands for the mean squared error between the original image and the generated image, which means that X_{ij} and Y_{ij} are the values of the pixels in the i-th row and the j-th column in the original image and the generated image, respectively.

Table 4.3: The performance evaluation of SR and Denoising technique.

Method	**Input-output**	**PSNR(dB)**
Super-Resolution	$\text{Image}_{12\times16}\rightarrow\text{Image}_{24\times32}$	32.62
	$\text{Image}_{6\times8}\rightarrow\text{Image}_{24\times32}$	20.47
Denoising	$\text{Image}_{24\times32}$	34.12
	$\text{Image}_{12\times16}$	30.52
	$\text{Image}_{6\times8}$	23.74

Where m represents the numbers of rows of pixels and n represents the number of columns of pixels. In general, the higher the MSE, the less similar the generated image is compared to the original, thus the PSNR decreases. In other words, a higher PSNR indicates a higher quality image.

Table 4.3 lists the result of SR and denoising. As can be seen, the PSNR of the 12×16 frames reaches 32.62 dB. As for 6×8 frames, the PSNR reaches 20.47 dB. The denoising result performs well for the 24×32 frames. The PSNR reaches 34.62 dB. This means that the denoised frames have good quality allowing for improving the predictions. As for the low resolution 12×16 and 6×8 frames, the PSNR has also been improved. However, it is not enough to generate good quality images as the HR24×32 ones.

4.6.2 Overall Classification Results

We use accuracy as metric for evaluating the efficiency of activity detection classification. Using the True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values, the accuracy is calculated based on the following formulas:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN}. \tag{4.12}$$

First we report the overall classification accuracy of various techniques. Table 4.4 show the classification accuracy achieved using the CNN. As evident, the individual application of SR, denoising, and CGAN has helped obtain better results. The ac-

Table 4.4: The overall activity classification results using CNN.

Method	**Image$_{6\times8}$**	**Image$_{12\times16}$**	**Image$_{24\times32}$**
Raw data	76.57%	88.22%	93.12%
Interpolation [202]	76.81%	88.56%	93.53%
SR	77.72%	89.24 %	–
Denoising	76.88%	88.30%	93.71%
CGAN+Raw data	–	–	95.24%
Denoising → SR	78.25%	89.31%	–
SR → Denoising	80.47%	91.72%	–
Denoising + CGAN	–	–	96.54%
Denoising → SR + CGAN	81.12%	92.66%	–
SR → Denoising + CGAN	83.58%	94.44%	–

Table 4.5: The overall activity classification results using CNN+LSTM.

S.No.	**Image$_{6\times8}$**	**Image$_{12\times16}$**	**Image$_{24\times32}$**
Raw data	78.32%	90.11%	95.73%
Interpolation [202]	78.68%	90.27%	96.08%
SR	79.07%	90.89%	–
Denoising	78.55%	90.33%	96.14%
CGAN+Raw data	–	–	96.42%
Denoising → SR	80.18%	92.38%	–
SR → Denoising	82.76%	92.91%	–
Denoising + CGAN	–	–	98.12%
Denoising → SR + CGAN	80.41%	93.43%	–
SR → Denoising + CGAN	84.43%	94.52%	–

Table 4.6: A comparison between the results achieved with our proposed approach and those achieved by employing some of the existing methods in the literature.

Approach	**Image$_{6\times8}$**	**Image$_{12\times16}$**	**Image$_{24\times32}$**
SVM [67]	61.45%	68.52%	88.16%
CNN [72]	67.11%	82.97%	90.14%
3D-CNN [168]	72.42%	90.89%	93.28%
CNN+LSTM$_{\text{SR → DE + CGAN}}$	**84.43%**	**94.52%**	–
CNN+LSTM$_{\text{DE + CGAN}}$	–	–	**98.12%**

curacy of classification using the raw data is high, it has been observed that the SR technique has further enhanced the accuracy of the result. The denoising technique, has also generated better results, though the accuracy of classification of the denoised frames is less than that of SR. Furthermore, the combined application of CGAN and raw data has noticeably improved the classification accuracy from 93.12% to 95.24%. Besides experimenting with each technique aside, we applied a combination of the preprocessing techniques on the low resolution data. As shown in Table 4.4, three specific combinations have been applied namely,

1. Super-Resolution → Denoising,
2. Denoising → SR,
3. Denoising & CGAN.

Clearly, the classification accuracy of the frames enhanced using all these combinations has improved further. We further combined the preprocessing techniques with the augmented data, wherein the derived results reflect the improved classification results, which is as high as that of the original HR(i.e., 24×32) frames. For the 6×8 low resolution data, the classification results reached 83.58%, while for the 12×16 low resolution data, it reached 94.44%. Remarkably, the maximum accuracy of classification of the original data with the HR (i.e., 24×32) has been obtained through the combined application of 'Denoising & CGAN', and reached 96.54%.

Further, Table 4.5 presents the results of sequential classification, wherein the output from the CNN is utilized as an input to the LSTM, with a time window size of five frames. Here, we observe that the sequential classification of the raw data and low resolution data has improved considerably. By applying the SR and denoising (both independently and collectively), the data classification results have further been improved. Also, using data augmentation techniques, we found that the sequential classification accuracy has increased. From Tables 4.5 and 4.4 we observed when we apply the interpolation technique proposed in [202] on the current data, the performance is not good, and is almost equal to that when we use the raw data. A very small improvement is observed.

For a fairer evaluation of our approach, we ran it against some existing approaches, namely ones that employ SVM, CNN, and 3D-CNN classifier. Table 4.6 compares the classification accuracy for these models with the highest accuracy of our approach. As can be seen from the table, our proposed approach outperforms the existing ones.

4.6.3 Activity Classification Results

Further experiments are run to evaluate the contribution of the different image enhancement techniques. We report activity classification accuracy using various combinations of these techniques. For simplification, and since we used different techniques, we use the following terminology for each of the resolutions or techniques used:

- The raw sensor data with various resolution are referred to as $R_{(24\times32)}$, $R_{(12\times16)}$, and $R_{(6\times8)}$.
- The SR technique applied to LR data are referred to as $SR_{(12\times16)}$, and $SR_{(6\times8)}$.
- The denoising technique applied to the HR and LR data are referred to as $DE_{(24\times32)}$, $DE_{(12\times16)}$, and $DE_{(6\times8)}$.
- The combination of raw data with CGAN techniques is referred to as $R + CG_{(24\times32)}$
- The denoising and CGAN techniques applied to HR data is referred to as $DE + CG_{(24\times32)}$
- The combination of denoising and SR techniques applied to LR data is referred to as $DE \rightarrow SR_{(12\times16)}$, $DE \rightarrow SR_{(6\times8)}$, $SR \rightarrow DE_{(12\times16)}$, and $SR \rightarrow DE_{(6\times8)}$.
- The combination of SR, denoising and CGAN techniques applied to LR data is referred to as $DE \rightarrow SR + CG_{(12\times16)}$, $DE \rightarrow SR + CG_{(6\times8)}$, $SR \rightarrow DE + CG_{(12\times16)}$, and $SR \rightarrow DE + CG_{(6\times8)}$.

Table 4.7: The results of activity classification using CNN on 6×8 data.

Method	**Walking**	**Standing**	**Sitting**	**Lying**	**Action change**	**Falling**
$R_{(6\times8)}$	80%	68%	77%	68%	57%	55%
$SR_{(6\times8)}$	86%	73%	86%	76%	60%	63%
$DE_{(6\times8)}$	85%	86%	82%	68%	62%	64%
$DE \rightarrow SR_{(6\times8)}$	86%	78%	86%	70%	70%	62%
$SR \rightarrow DE_{(6\times8)}$	84%	82%	84%	72%	73%	71%
$DE \rightarrow SR + CG_{(6\times8)}$	82%	80%	79%	75%	78%	70%
$SR \rightarrow DE + CG_{(6\times8)}$	80%	84%	83%	81%	80%	68%

Table 4.8: The results of activity classification using CNN on 12×16 data.

Method	**Walking**	**Standing**	**Sitting**	**Lying**	**Action change**	**Falling**
$R_{(12\times16)}$	80%	88%	86%	75%	82%	81%
$SR_{(12\times16)}$	84%	86%	81%	85%	84%	83%
$DE_{(12\times16)}$	83%	85%	86%	85%	84%	79%
$DE \rightarrow SR_{(12\times16)}$	84%	85%	89%	90%	82%	84%
$SR \rightarrow DE_{(12\times16)}$	90%	92%	88%	91%	90%	94%
$DE \rightarrow SR{+}CG_{(12\times16)}$	89%	94%	90%	85%	84%	88%
$SR \rightarrow DE{+}CG_{(12\times16)}$	91%	92%	89%	92%	93%	91%

Table 4.9: The results of activity classification using CNN on 24×32 data.

Method	**Walking**	**Standing**	**Sitting**	**Lying**	**Action change**	**Falling**
$R_{(24\times32)}$	88%	90%	87%	93%	90%	91%
$DE_{(24\times32)}$	92%	84%	92%	90%	91%	89%
$R + CG_{(24\times32)}$	90%	95%	90%	94%	92%	92%
$DE + CG_{(24\times32)}$	96%	92%	95%	94%	87%	93%

The results of activity classification accuracy of CNN using LR 6×8 data are shown in Table 4.7. We can see that each technique improved the activity classification performance. The walking and sitting, for example, present high accuracy in the $DE \rightarrow SR_{(6\times8)}$ technique, reaching both 86%. Also, the falling accuracy is 71% in $SR \rightarrow DE_{(6\times8)}$. Data augmentation aids in the improvement of activity detection in $SR \rightarrow DE + CG_{(6\times8)}$ for other activities, like standing, lying and action change reaches high accuracy of 84%, 81%, and 80%, respectively.

The results of activity classification accuracy of CNN using LR 12×16 are shown in Table 4.8. Here, for the particular case of the falling activity, $SR \rightarrow DE_{(12\times16)}$ technique reaches a high accuracy of 94%. Similarly the standing accuracy is 94% in $DE \rightarrow SR + CG_{(12\times16)}$ technique. The performance of detection of other activities improved using the $SR \rightarrow DE + CG_{(12\times16)}$ technique, reaching a maximum of 93% accuracy in the action change activity, 92% accuracy in standing and lying.

We infer from these CNN classification results that performing SR followed by denoising and then adding CGAN data improves performance.

Table 4.9 shows the results of the HR data classification using CNN. Here, $DE + CG_{(24\times32)}$ technique performs well for the majority of the activities. Walking reaches an accuracy of 96%, sitting reaches an accuracy of 95%, and lying reaches an accuracy of 94%. The performance boost provided by denoising, which creates a clear image, aids in detecting activity.

Table 4.10 shows the activity classification results using CNN+LSTM on LR 6×8 data. Here, again we can see that each technique improves the activity classification performance. For example, $SR \rightarrow DE_{(6\times8)}$ has a high lying and action change accuracy of 82%. The classification accuracy of walking, standing, and sitting activities in $SR \rightarrow DE + CG_{(6\times8)}$ is 84%, 82%, and 84%, respectively. This is thanks to the image qualify improvement after applying SR, then DE.

Table 4.11 shows the activity classification results using CNN+LSTM on LR 12×16 data. Here, sitting and lying activities, after applying $DE \rightarrow SR + CG_{(12\times16)}$ reaches a high accuracy of 93% and 94%, respectively. Other activities like walking and action changes reaches 93% accuracy in $SR_{(12\times16)}$ and $SR \rightarrow DE + CG_{(12\times16)}$

techniques, respectively. This is thanks to the fact that 12×16 resolution data contains significantly more information than 6×8 resolution data. By applying the denoising technique after SR, images are smoothed and enhanced making it easier to detect sitting and lying activities.

The results of activity classification using CNN+LSTM applied on HR data is shown in Table 4.12. In most activities, the $DE + CG_{(24\times32)}$ technique performs well. Walking has a highest classification accuracy of 96%, action change has that of 97%, lying has that of 94%, and falling has that of 96%. Denoising further improves the performance by making the image clearer, which makes it easier to recognize the activity.

4.6.4 Neural Network Quantization

As previously stated, we used quantization on our neural network because one of our primary objectives is to have the proposed approach running on low-powered devices. Although the process of quantization, generally speaking, reduces the accuracy, it can still be used given a flexibility for a trade-off between performance and optimization. In a first step, we compare the classification accuracy of our models with and without quantization, when using the raw data (i.e., no image enhancement or data augmentation is used). Table 4.13 compares the performance of classification of such raw data with and without quantization. As can be seen, the accuracy is reduced, but not significantly. The accuracy drops range from 0.06% for images of resolution 12×16 to 2.09% for images of resolution 6×8. However, to recall, our main goal is to optimize the proposed deployment approach both in terms of performance and complexity. Given the different techniques proposed in this work, we compare the results of the best performing techniques with and without quantization in Table 4.14. As can be seen, the table shows a degradation in accuracy to some extent compared to when quantization is not used. However, that does not negate the many advantages of quantization; model sizes are reduced, and inference times are reduced to the point where they are more beneficial in low-end devices. All these results are generated using 8-bit integer data on our computer.

4.7 Comparative Study Analysis

Table 4.15 presents a comparative illustration of the results derived through the Chapter 3 and Chapter 4 of the current thesis. Evidently, the setup for conducting the two experiments is considerably varying. Chapter 3, two sensors have been used for activity detection (placed on ceiling and wall), whereas in Chapter 4 only one sensor has been used (played in ceiling). Conducted in two rooms, no preprocessing was done on the date derived through Chapter 3. On the other hand, three different rooms were utilized for the experiment in Chapter 4, and the derived data was preprocessed using super resolution and denoising techniques. This helps to enhance the quality of the data. Further, for the classification of results, we used the same neural network architecture composed of CNN and LSTM in both approaches. However, CGAN data are used in the training only in Chapter 4. Remarkably, the results derived through both the experiments showcased high accuracy of 97% and 98% respectively. Even more so, the approach using one sensor has delivered good results due to various deep learning approaches, which is unprecedented in the field of activity detection. Notably, the use of one sensor not only minimizes the cost, but also reduces the inference time and the model size, allowing it to run on low powered devices.

We tried to reproduce the existing approaches, run them on our data set, and compared their performance with our proposed approaches. Table 4.16 shows the comparison of the performance. It clearly shows that the our proposed approach outperforms the existing approaches in terms of inference time and classification accuracy. If we look at Table 4.16, we can see that our experiment with two sensors has marked a computation time of 2.88 seconds, whereas that with one sensor approach reaches 0.43 seconds.

Table 4.10: The results of activity classification using CNN+LSTM on 6×8 data.

Method	**Walking**	**Standing**	**Sitting**	**Lying**	**Action change**	**Falling**
$R_{(6\times8)}$	75%	78%	74%	77%	70%	74%
$SR_{(6\times8)}$	77%	74%	73%	78%	73%	76%
$DE_{(6\times8)}$	76%	72%	78%	70%	75%	72%
$DE \rightarrow SR_{(6\times8)}$	81%	78%	78%	75%	77%	73%
$SR \rightarrow DE_{(6\times8)}$	78%	81%	73%	82%	82%	75%
$DE \rightarrow SR + CG_{(6\times8)}$	80%	74%	70%	76%	78%	75%
$SR \rightarrow DE + CG_{(6\times8)}$	84%	82%	84%	78%	81%	80%

Table 4.11: The results of activity classification using CNN+LSTM on 12×16 data.

Method	**Walking**	**Standing**	**Sitting**	**Lying**	**Action change**	**Falling**
$R_{(12\times16)}$	88%	90%	76%	77%	80%	82%
$SR_{(12\times16)}$	85%	88%	82%	90%	86%	79%
$DE_{(12\times16)}$	82%	86%	78%	89%	90%	83%
$DE \rightarrow SR_{(12\times16)}$	92%	84%	77%	89%	91%	86%
$SR \rightarrow DE_{(12\times16)}$	93%	90%	88%	84%	91%	88%
$DE \rightarrow SR{+}CG_{(12\times16)}$	87%	90%	93%	94%	82%	90%
$SR \rightarrow DE{+}CG_{(12\times16)}$	90%	92%	90%	86%	93%	92%

Table 4.12: The results of activity classification using CNN+LSTM on 24×32 data.

Method	**Walking**	**Standing**	**Sitting**	**Lying**	**Action change**	**Falling**
$R_{(24\times32)}$	92%	91%	93%	90%	94%	89%
$DE_{(24\times32)}$	93%	95%	96%	91%	94%	92%
$R + CG_{(24\times32)}$	95%	94%	95%	93%	92%	90%
$DE + CG_{(24\times32)}$	96%	94%	93%	96%	97%	96%

Table 4.13: The performance comparison of raw data with quantization aware training.

Resolution	**With quantization**	**Accuracy**	**100 epochs training time (s)**	**Inference time (ms)**	**Model size (MB)**
6×8	Yes	76.23%	17	0.003	0.3
	No	78.32%	54	0.048	1.4
12×16	Yes	90.05%	36	0.007	0.8
	No	90.11%	88	0.078	2.4
24×32	Yes	94.20%	44	0.009	1.1
	No	95.73%	132	0.093	3.2

Table 4.14: A comparison between the performance of classification with and without quantization applied to the preprocessed and enhanced images using the techniques proposed above.

Resolution	**With quantization**	**Accuracy**	**100 epochs training time (s)**	**Inference time (ms)**	**Model size (MB)**
$6\times8_{\text{SR} \rightarrow \text{DE + CGAN}}$	Yes	82.27%	145	0.38	4.18
	No	84.43%	321	2.57	10.20
$12\times16_{\text{SR} \rightarrow \text{DE + CGAN}}$	Yes	93.18%	164	0.60	5.43
	No	94.52%	352	3.21	14.68
$24\times32_{\text{DE + CGAN}}$	Yes	97.53%	136	0.43	4.37
	No	98.12%	291	2.82	11.20

Table 4.15: Comparison of Chapter 3 and Chapter 4

Comparison	**Chapter 3**	**Chapter 4**
No. of Sensor	Two	One
Position	Ceiling and Wall	Ceiling
Preprocessing	No Preprocessing	Super Resolution and Denoising
Experiment	2 different rooms	3 different rooms
Classification	CNN+LSTM	CNN+LSTM(CGAN data is used for Training)
Accuracy	97%	98%

Table 4.16: Comparison of existing work with the proposed approaches.

Comparison	**Method**	**No. of Sensors**	**Position**	**Pre-processing**	**Data Augmentation**	**Accuracy**	**Inference time (ms)**
Conventional Machine Learning	SVM [15]	1	Ceiling	Thresholding	No	72%	2.17
	k-NN [55]	1	Ceiling	Thresholding	No	84%	2.20
	SVM [67]	2	Ceiling	Based on Motion Detection	No	90%	2.86
Deep Learning	CNN [72]	1	Ceiling and Wall	Fuzzy set representation of data	Traditional approach (flipping rotating the image)	90%	3.17
	3D-CNN [168]	1	Ceiling	Gaussian filter	No	93%	4.12
Chapter 3	CNN+LSTM	2	Ceiling and Wall	No Pre-processing	No	97%	2.88
Chapter 4	CNN+LSTM	1	Ceiling	SR and Denoising	CGAN	98%	0.43

Chapter 5

Conclusion and Future Work

5.1 Conclusions

To conclude the thesis overall, we propose a lightweight DL model for activity classification that is robust to environmental changes. Being lightweight, such a model can run on devices with very low computation capabilities, making it a base for a cheap solution for activity detection. The blurriness and noise present in the IR captured frames, due to the sensor characteristics the imprecision in the sensor lead to a noticeable drop in performance in conventional methods. Our proposed neural network architecture manages to address this issue by exploiting the temporal changes in the frames to identify the activities accurately. We identify the activity using a time window of less than 1 second. Despite the smaller time window, we have remarkably enhanced the classification accuracy in comparison to conventional works, which require a larger time window. LR sensors are always preferred over HR ones if they provide similar performance. This is thanks to their lower risk of privacy invasion and cheaper cost. We demonstrate that it is possible to use the LR data to achieve classification performance that is nearly identical to that of the classification of the HR data, namely 24×32, by using deep learning techniques such as Super-resolution, denoising, and CGAN.

In this thesis, first we proposed an activity detection technique using two wide-angle low-resolution IR array sensors. The data collected by the sensors are classified

using a hybrid DL model. The hybrid DL model is designed based on the Convolution-LSTM. We used two sensors, one placed on the wall and the other placed on the ceiling. This activity detection system involves two phases. In the first phase we classify the wall sensor data and ceiling sensor data using CNN and achieve a classification accuracy of 0.93 and 0.94, respectively. To improve this further, we combined both the sensor data and performed classification using CNN and got an improved accuracy of 0.96. In the second phase, the output of the CNN is passed to an LSTM to achieve better performance. The classification using the ceiling sensor data reaches 0.96 accuracy, whereas that using wall sensor data reaches 0.95 accuracy. When we combine both the wall sensor data and ceiling sensor data, the classification accuracy reaches 0.97. We run some of the existing conventional approaches on our data set and compared the results. Based on these, we can conclude that by combining the data collected by the sensor placed on the ceiling and that placed on the wall, and using CNN and LSTM, we get the highest classification accuracy which is 0.97.

To further enhance the detection and optimize the AD system, we used one IR array placed on the ceiling, and conducted the various experiments in which we collected data under different conditions for a continuous period of time with different resolutions (i.e., 24×32, 12×16, and 6×8) To identify the activity of the participants, we ran a classification task that takes the frames generated by the sensor as the input and predicts the activity. To further enhance the classification, we applied three advanced DL techniques: SR, denoising, and CGAN. Herein, the key purpose was to enhance the classification accuracy of the low-resolution data. Through the results, we observed that the application of these techniques has helped improve the classification accuracy of low-resolution images from 78.32% to 84.43% (6×8 resolution) and from 90.11% to 94.54% (12×16 resolution). We optimize the classification model to run on low-powered devices. We used quantization on our neural network. It reduces the model size and the running time of the prediction, trade off with classification accuracy.

5.2 Future Work

Focused on enhancing the activity detection, this research has used two-sensors for activity detection. Further, to proving a cheaper solution with similar performance, only one sensor was used, through which the model has been observed to be highly optimized. It has been noted that in such a manner, the activity detection system can also be run through low-power devices. Notably, by using low-resolution data, nearly high-resolution data performance has been achieved in this research. Based on our obtained results presented in this work, future studies include applying SR and denisoing techniques on two sensor placed ceiling and wall, focusing on stereoscopic vision using two sensors, and trying to build a 3D reconstruction of person to identify his/her activities. In that way, we can improve the quality of data and remove noise. We can also determine the height and depth of a person while performing any activity and can effectively predict the fall with high precision. Moreover, this research approach has been robust from environment perspective, such that it can work even in unseen conditions.

Towards the end, it is important to highlight that this research is subject to certain limitations. Firstly, the current research only considers one person and not more. Also, if there is any heat emitting object, the activity detection system proposed in this research may not deliver accurate results. Due to the optimization of neural networks, the classification accuracy and the performance of activity detection is also being reduced. Future studies should therefore address these research limitations.

Bibliography

[1] S. B. M. of Internal Affairs and C. Japan, "Statistical handbook of japan 2021," *JAPAN STATISTICAL YEARBOOK*, 2021.

[2] "Annual report on the ageing society." [Online]. Available: https://www8.cao.go.jp/kourei/english/annualreport/2019/pdf/2019.pdf

[3] T. Le Nguyen and T. T. H. Do, "Artificial intelligence in healthcare: A new technology benefit for both patients and doctors," in *2019 Portland International Conference on Management of Engineering and Technology (PICMET)*. IEEE, 2019, pp. 1–15.

[4] Y. Song and T. J. van der Cammen, "Electronic assistive technology for community-dwelling solo-living older adults: A systematic review," *Maturitas*, vol. 125, pp. 50–56, 2019.

[5] F. G. Miskelly, "Assistive technology in elderly care," *Age and ageing*, vol. 30, no. 6, pp. 455–458, 2001.

[6] S. A. Zwijsen, A. R. Niemeijer, and C. M. Hertogh, "Ethics of using assistive technology in the care for community-dwelling elderly people: An overview of the literature," *Aging & mental health*, vol. 15, no. 4, pp. 419–427, 2011.

[7] D. Moreira, M. Barandas, T. Rocha, P. Alves, R. Santos, R. Leonardo, P. Vieira, and H. Gamboa, "Human activity recognition for indoor localization using smartphone inertial sensors," *Sensors*, vol. 21, no. 18, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/18/6316

[8] F. Shahmohammadi, A. Hosseini, C. E. King, and M. Sarrafzadeh, "Smartwatch based activity recognition using active learning," in *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 2017, pp. 321–329.

[9] A. A. Sukor, A. Zakaria, and N. A. Rahim, "Activity recognition using accelerometer sensor and machine learning classifiers," in *Proc. IEEE 14th Int. Colloq. on Signal Processing and Its Applications (CSPA)*. IEEE, 2018, pp. 233–238.

[10] M. L. Gavrilova, Y. Wang, F. Ahmed, and P. P. Paul, "Kinect sensor gesture and activity recognition: New applications for consumer cognitive systems," *IEEE Consumer Electronics Magazine*, vol. 7, no. 1, pp. 88–94, 2017.

[11] Y. Hino, J. Hong, and T. Ohtsuki, "Activity recognition using array antenna," in *Proc. IEEE Int. Conf. Communications (ICC)*. IEEE, 2015, pp. 507–511.

[12] J. Hong, S. Tomii, and T. Ohtsuki, "Cooperative fall detection using doppler radar and array sensor," in *Proc. IEEE 24th Annu. Int. Symp. on Personal, Indoor, and Mobile Radio Commun. (PIMRC)*. IEEE, 2013, pp. 3492–3496.

[13] M. Bouazizi, C. Ye, and T. Ohtsuki, "2d lidar-based approach for activity identification and fall detection," *IEEE Internet of Things Journal*, pp. 1–1, 2021.

[14] T. Nakamura, M. Bouazizi, K. Yamamoto, and T. Ohtsuki, "Wi-fi-csi-based fall detection by spectrogram analysis with cnn," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2020, pp. 1–6.

[15] S. Mashiyama, J. Hong, and T. Ohtsuki, "Activity recognition using low resolution infrared array sensor," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2015, pp. 495–500.

[16] Z. Yang, Z. Wang, J. Zhang, C. Huang, and Q. Zhang, "Wearables can afford: Light-weight indoor positioning with visible light," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, 2015, pp. 317–330.

[17] Y. H. Lee and G. Medioni, "Rgb-d camera based wearable navigation system for the visually impaired," *Computer vision and Image understanding*, vol. 149, pp. 3–20, 2016.

[18] "Apple pay." [Online]. Available: https://www.apple.com/apple-pay/

[19] "visa." [Online]. Available: https://usa.visa.com/pay-with-visa/find-card/buy-gift-card

[20] M. Vidal, J. Turner, A. Bulling, and H. Gellersen, "Wearable eye tracking for mental health monitoring," *Computer Communications*, vol. 35, no. 11, pp. 1306–1311, 2012.

[21] J. Wijsman, B. Grundlehner, H. Liu, H. Hermens, and J. Penders, "Towards mental stress detection using wearable physiological sensors," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 1798–1801.

[22] D. Anzaldo, "Wearable sports technology-market landscape and compute soc trends," in *2015 International SoC Design Conference (ISOCC)*. IEEE, 2015, pp. 217–218.

[23] [Online]. Available: https://www.forbes.com/

[24] "Wearables market sees first decline at beginning of 2022 as demand normalizes, according to idc." [Online]. Available: https://www.idc.com/getdoc.jsp?containerId=prUS49250022

[25] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, "A review of wearable sensors and systems with application in rehabilitation," *Journal of neuroengineering and rehabilitation*, vol. 9, no. 1, pp. 1–17, 2012.

[26] M. Al-Khafajiy, T. Baker, C. Chalmers, M. Asim, H. Kolivand, M. Fahim, and A. Waraich, "Remote health monitoring of elderly through wearable sensors," *Multimedia Tools and Applications*, vol. 78, no. 17, pp. 24 681–24 706, 2019.

[27] J. Liu, J. Sohn, and S. Kim, "Classification of daily activities for the elderly using wearable sensors," *Journal of healthcare engineering*, vol. 2017, 2017.

[28] Z. Wang, Z. Yang, and T. Dong, "A review of wearable technologies for elderly care that can accurately track indoor position, recognize physical activities and monitor vital signs in real time," *Sensors*, vol. 17, no. 2, p. 341, 2017.

[29] J. Heikenfeld, A. Jajack, J. Rogers, P. Gutruf, L. Tian, T. Pan, R. Li, M. Khine, J. Kim, and J. Wang, "Wearable sensors: modalities, challenges, and prospects," *Lab on a Chip*, vol. 18, no. 2, pp. 217–248, 2018.

[30] A. Kamišalić, I. Fister Jr, M. Turkanović, and S. Karakatič, "Sensors and functionalities of non-invasive wrist-wearable devices: A review," *Sensors*, vol. 18, no. 6, p. 1714, 2018.

[31] J. Huberty, D. K. Ehlers, J. Kurka, B. Ainsworth, and M. Buman, "Feasibility of three wearable sensors for 24 hour monitoring in middle-aged women," *BMC women's health*, vol. 15, no. 1, pp. 1–9, 2015.

[32] M. H. Bahari, L. K. Hamaidi, M. Muma, J. Plata-Chaves, M. Moonen, A. M. Zoubir, and A. Bertrand, "Distributed multi-speaker voice activity detection for wireless acoustic sensor networks," *arXiv preprint arXiv:1703.05782*, 2017.

[33] Y. Li, Z. Zeng, M. Popescu, and K. Ho, "Acoustic fall detection using a circular microphone array," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, 2010, pp. 2242–2245.

[34] Z. Fu, T. Delbruck, P. Lichtsteiner, and E. Culurciello, "An address-event fall detector for assisted living applications," *IEEE transactions on biomedical circuits and systems*, vol. 2, no. 2, pp. 88–96, 2008.

[35] T. Hasija, M. Gölz, M. Muma, P. J. Schreier, and A. M. Zoubir, "Source enumeration and robust voice activity detection in wireless acoustic sensor networks," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2019, pp. 1257–1261.

[36] V. Vishwakarma, C. Mandal, and S. Sural, "Automatic detection of human fall in video," in *International conference on pattern recognition and machine intelligence*. Springer, 2007, pp. 616–623.

[37] M. Zia Uddin, W. Khaksar, and J. Torresen, "A thermal camera-based activity recognition using discriminant skeleton features and rnn," in *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, vol. 1, 2019, pp. 777–782.

[38] L. Becker, "Influence of ir sensor technology on the military and civil defense," in *Quantum Sensing and Nanophotonic Devices III*, vol. 6127. SPIE, 2006, pp. 180–194.

[39] S. Park, H. T. Kim, S. Lee, H. Joo, and H. Kim, "Survey on anti-drone systems: Components, designs, and challenges," *IEEE Access*, vol. 9, pp. 42 635–42 659, 2021.

[40] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Robust video surveillance for fall detection based on human shape deformation," *IEEE Transactions on circuits and systems for video Technology*, vol. 21, no. 5, pp. 611–622, 2011.

[41] J. C.-W. Cheung, E. W.-C. Tam, A. H.-Y. Mak, T. T.-C. Chan, and Y.-P. Zheng, "A night-time monitoring system (enightlog) to prevent elderly wandering in hostels: a three-month field study," *International journal of environmental research and public health*, vol. 19, no. 4, p. 2103, 2022.

[42] S. S. Torkestani, S. Sahuguede, A. Julien-Vergonjanne, J. Cances, and J.-C. Daviet, "Infrared communication technology applied to indoor mobile healthcare monitoring system," *International Journal of E-Health and Medical Communications (IJEHMC)*, vol. 3, no. 3, pp. 1–11, 2012.

[43] M. B. Coskun and M. Rais-Zadeh, "Thermal infrared detector sparse array for nasa planetary applications," in *Proc. 5th IEEE Conf. Electron Devices Technology & Manufacturing (EDTM)*, 2021, pp. 1–3.

[44] E. Josse, A. Nerborg, K. Hernandez-Diaz, and F. Alonso-Fernandez, "In-bed person monitoring using thermal infrared sensors," in *Proc. 16th Conf. Computer Science and Intelligence Systems (FedCSIS)*, 2021, pp. 121–125.

[45] D. L. Luu, C. Lupu, I. Cristian *et al.*, "Speed control and spacing control for autonomous mobile robot platform equipped with infrared sensors," in *Proc. 16th Int. Conf. on Engineering of Modern Electric Systems (EMES)*, 2021, pp. 1–4.

[46] S. Lee, K. N. Ha, and K. C. Lee, "A pyroelectric infrared sensor-based indoor location-aware system for the smart home," *IEEE Transactions on Consumer Electronics*, vol. 52, no. 4, pp. 1311–1317, 2006.

[47] Q. Liang, L. Yu, X. Zhai, Z. Wan, and H. Nie, "Activity recognition based on thermopile imaging array sensor," in *Proc. IEEE Int. Conf. on Electro/Inf. Technol.(EIT)*, 2018, pp. 0770–0773.

[48] K. A. Muthukumar, M. Bouazizi, and T. Ohtsuki, "Activity detection using wide angle low-resolution infrared array sensors," in *Proc. The Inst. of Electronics, Inf. and Commun. Engineers (IEICE) Conf. Archives.*, 2020, pp. BS–8–1.

[49] S. O. Al-Jazzar, S. A. Aldalahmeh, D. McLernon, and S. A. R. Zaidi, "Intruder localization and tracking using two pyroelectric infrared sensors," *IEEE Sensors Journal*, vol. 20, no. 11, pp. 6075–6082, 2020.

[50] M. Bouazizi and T. Ohtsuki, "An infrared array sensor-based method for localizing and counting people for health care and monitoring," in *Proc. 42nd Annu. Int. Conf. of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 4151–4155.

[51] T. Yang, P. Guo, W. Liu, X. Liu, and T. Hao, "Enhancing pir-based multi-person localization through combining deep learning with domain knowledge," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 4874–4886, 2021.

[52] D. B. Sam, S. V. Peri, M. N. Sundararaman, A. Kamath, and R. V. Babu, "Locate, size, and count: Accurately resolving people in dense crowds via detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 8, pp. 2739–2751, 2021.

[53] C. Wu, F. Zhang, B. Wang, and K. J. Ray Liu, "Proc. mmtrack: Passive multi-person localization using commodity millimeter wave radio," in *IEEE INFOCOM 2020 - IEEE Conf. on Computer Commun.*, 2020, pp. 2400–2409.

[54] M. Bouazizi, C. Ye, and T. Ohtsuki, "Low-resolution infrared array sensor for counting and localizing people indoors: When low end technology meets cutting edge deep learning techniques," *Information*, vol. 13, no. 3, 2022. [Online]. Available: https://www.mdpi.com/2078-2489/13/3/132

[55] S. Mashiyama, J. Hong, and T. Ohtsuki, "A fall detection system using low resolution infrared array sensor," in *Proc. IEEE Annu. Int. Symp. on Personal, Indoor, and Mobile Radio Commun. (PIMRC)*, 2014, pp. 2109–2113.

[56] X. Luo, Q. Guan, H. Tan, L. Gao, Z. Wang, and X. Luo, "Simultaneous indoor tracking and activity recognition using pyroelectric infrared sensors," *Sensors*, vol. 17, no. 8, 2017. [Online]. Available: https://www.mdpi.com/1424-8220/17/8/1738

[57] N. Gu, B. Yang, and T. Zhang, "Dynamic fuzzy background removal for indoor human target perception based on thermopile array sensor," *IEEE Sensors Journal*, vol. 20, no. 1, pp. 67–76, 2019.

[58] R. Tang, T. Zhang, Y. Chen, H. Liang, B. Li, and Z. Zhou, "Infrared thermography approach for effective shielding area of field smoke based on background subtraction and transmittance interpolation," *Sensors*, vol. 18, no. 5, p. 1450, 2018.

[59] E. S. Jeon, J.-S. Choi, J. H. Lee, K. Y. Shin, Y. G. Kim, T. T. Le, and K. R. Park, "Human detection based on the generation of a background image by using a far-infrared light camera," *Sensors*, vol. 15, no. 3, pp. 6763–6788, 2015.

[60] E. Goubet, J. Katz, and F. Porikli, "Pedestrian tracking using thermal infrared imaging," in *Infrared technology and applications XXXII*, vol. 6206. SPIE, 2006, pp. 797–808.

[61] A. Naser, A. Lotfi, and J. Zhong, "Towards human distance estimation using a thermal sensor array," *Neural Computing and Applications*, pp. 1–11, June 2021.

[62] S. Okuda, S. Kaneda, and H. Haga, "Human position/height detection using analog type pyroelectric sensors," in *International Conference on Embedded and Ubiquitous Computing*. Springer, 2005, pp. 306–315.

[63] X. Zhang, H. Seki, and M. Hikizu, "Detection of human position and motion by thermopile infrared sensor," *International Journal of Automation Technology*, vol. 9, no. 5, pp. 580–587, 2015.

[64] S. Parnin and M. Rahman, "Human location detection system using microelectromechanical sensor for intelligent fan," in *IOP Conference Series: Materials Science and Engineering*, vol. 184, no. 1. IOP Publishing, 2017, p. 012042.

[65] T. Hosokawa and M. Kudo, "Person tracking with infrared sensors," in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. Springer, 2005, pp. 682–688.

[66] K. A. Kumar and O. Dhadge, "A novel infrared (ir) based sensor system for human presence detection in targeted locations." *International Journal of Computer Network & Information Security*, vol. 10, no. 12, 2018.

[67] K. Kobayashi, T. Ohtsuki, and K. Toyoda, "Human activity recognition by infrared sensor arrays considering positional relation between user and sensors," in *Proc. Smart City Based Ambient Intelliegnce*, 2018, pp. 1–6.

[68] X. Fan, H. Zhang, C. Leung, and Z. Shen, "Robust unobtrusive fall detection using infrared array sensors," in *Proc. IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2017, pp. 194–199.

[69] Y. Taniguchi, H. Nakajima, N. Tsuchiya, J. Tanaka, F. Aita, and Y. Hata, in *Proc. Int. Conf. on Soft Computing and Intelligent Systems (SCIS)*, 2014, pp. 673–678.

[70] C. Zhong, W. W. Ng, S. Zhang, C. D. Nugent, C. Shewell, and J. Medina-Quero, "Multi-occupancy fall detection using non-invasive thermal vision sensor," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 5377–5388, 2020.

[71] M. Burns, F. Cruciani, P. Morrow, C. Nugent, and S. McClean, "Using convolutional neural networks with multiple thermal sensors for unobtrusive pose recognition," *Sensors*, vol. 20, no. 23, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/23/6932

[72] M. Ángel López-Medina, M. Espinilla, C. Nugent, and J. M. Quero, "Evaluation of convolutional neural networks for the classification of falls from heterogeneous thermal vision sensors," *International Journal of Distributed Sensor Networks*, vol. 16, no. 5, 2020.

[73] D. Rand, J. J. Eng, P.-F. Tang, J.-S. Jeng, and C. Hung, "How active are people with stroke? use of accelerometers to assess physical activity," *Stroke*, vol. 40, no. 1, pp. 163–168, 2009.

[74] D. S. Ward, K. R. Evenson, A. Vaughn, A. B. Rodgers, and R. P. Troiano, "Accelerometer use in physical activity: best practices and research recommendations." *Medicine and science in sports and exercise*, vol. 37, no. 11 Suppl, pp. S582–8, 2005.

[75] J.-S. Lee and H.-H. Tseng, "Development of an enhanced threshold-based fall detection system using smartphones with built-in accelerometers," *IEEE Sensors Journal*, vol. 19, no. 18, pp. 8293–8302, 2019.

[76] G. L. Santos, P. T. Endo, K. H. d. C. Monteiro, E. d. S. Rocha, I. Silva, and T. Lynn, "Accelerometer-based human fall detection using convolutional neural networks," *Sensors*, vol. 19, no. 7, p. 1644, 2019.

[77] T.-L. Le, J. Morel *et al.*, "An analysis on human fall detection using skeleton from microsoft kinect," in *2014 IEEE Fifth International Conference on Communications and Electronics (ICCE)*. IEEE, 2014, pp. 484–489.

[78] S. Ranakoti, S. Arora, S. Chaudhary, S. Beetan, A. S. Sandhu, P. Khandnor, and P. Saini, "Human fall detection system over imu sensors using triaxial accelerometer," in *Computational Intelligence: Theories, Applications and Future Directions-Volume I*. Springer, 2019, pp. 495–507.

[79] X. Sun, Z. Lu, W. Hu, and G. Cao, "Symdetector: detecting sound-related respiratory symptoms using smartphones," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015, pp. 97–108.

[80] X. Sun, Z. Lu, X. Zhang, M. Salathé, and G. Cao, "Infectious disease containment based on a wireless sensor system," *Ieee Access*, vol. 4, pp. 1558–1569, 2016.

[81] X. Sun, L. Qiu, Y. Wu, Y. Tang, and G. Cao, "Sleepmonitor: Monitoring respiratory rate and body position during sleep using smartwatch," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 1, no. 3, pp. 1–22, 2017.

[82] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," in *International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)*. IEEE, 2006, pp. 4–pp.

[83] G. M. Weiss, J. L. Timko, C. M. Gallagher, K. Yoneda, and A. J. Schreiber, "Smartwatch-based activity recognition: A machine learning approach," in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, 2016, pp. 426–429.

[84] S. Mekruksavanich, N. Hnoohom, and A. Jitpattanakul, "Smartwatch-based sitting detection with human activity recognition for office workers syndrome," in *2018 International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON)*. IEEE, 2018, pp. 160–164.

[85] S. Mekruksavanich, A. Jitpattanakul, P. Youplao, and P. Yupapin, "Enhanced hand-oriented activity recognition based on smartwatch sensor data using lstms," *Symmetry*, vol. 12, no. 9, p. 1570, 2020.

[86] C. Dobbins and R. Rawassizadeh, "Towards clustering of mobile and smartwatch accelerometer data for physical activity recognition," in *Informatics*, vol. 5, no. 2. MDPI, 2018, p. 29.

[87] S. Al-Janabi and A. H. Salman, "Sensitive integration of multilevel optimization model in human activity recognition for smartphone and smartwatch applications," *Big data mining and analytics*, vol. 4, no. 2, pp. 124–138, 2021.

[88] A. Mannini and A. M. Sabatini, "Machine learning methods for classifying human physical activity from on-body accelerometers," *Sensors*, vol. 10, no. 2, pp. 1154–1175, 2010.

[89] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," in *International workshop on ambient assisted living*. Springer, 2012, pp. 216–223.

[90] S. Balli, E. A. Sağbaş, and M. Peker, "Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm," *Measurement and Control*, vol. 52, no. 1-2, pp. 37–45, 2019.

[91] N. M. Fung, J. W. S. Ann, Y. H. Tung, C. S. Kheau, and A. Chekima, "Elderly fall detection and location tracking system using heterogeneous wireless networks," in *2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*. IEEE, 2019, pp. 44–49.

[92] Y. Wang, K. Wu, and L. M. Ni, "Wifall: Device-free fall detection by wireless networks," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 581–594, 2016.

[93] S. P. Rana, M. Dey, M. Ghavami, and S. Dudley, "Signature inspired home environments monitoring system using ir-uwb technology," *Sensors*, vol. 19, no. 2, p. 385, 2019.

[94] H. Yoshino, V. G. Moshnyaga, and K. Hashimoto, "Fall detection on a single doppler radar sensor by using convolutional neural networks," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE, 2019, pp. 2889–2892.

[95] B. Y. Su, K. Ho, M. J. Rantz, and M. Skubic, "Doppler radar fall activity detection using the wavelet transform," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 3, pp. 865–875, 2014.

[96] H. Sadreazami, M. Bolic, and S. Rajan, "Capsfall: Fall detection using ultra-wideband radar and capsule network," *IEEE Access*, vol. 7, pp. 55 336–55 343, 2019.

[97] H. Sadreazami and M. Bolic, "Fall detection using standoff radar-based sensing and deep convolutional neural network," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 1, pp. 197–201, 2019.

[98] C. Ding, Y. Zou, L. Sun, H. Hong, X. Zhu, and C. Li, "Fall detection with multi-domain features by a portable fmcw radar," in *2019 IEEE MTT-S International Wireless Symposium (IWS)*. IEEE, 2019, pp. 1–3.

[99] B. Erol and M. G. Amin, "Radar data cube processing for human activity recognition using multisubspace learning," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 6, pp. 3617–3628, 2019.

[100] B. Jokanovic, M. Amin, and F. Ahmad, "Radar fall motion detection using deep learning," in *2016 IEEE radar conference (RadarConf)*. IEEE, 2016, pp. 1–6.

[101] C. Álvarez-Aparicio, Á. M. Guerrero-Higueras, F. J. Rodríguez-Lera, J. Ginés Clavero, F. Martín Rico, and V. Matellán, "People detection and tracking using lidar sensors," *Robotics*, vol. 8, no. 3, p. 75, 2019.

[102] Á. M. Guerrero-Higueras, C. Álvarez-Aparicio, M. C. Calvo Olivera, F. J. Rodríguez-Lera, C. Fernández-Llamas, F. M. Rico, and V. Matellán, "Tracking people in a mobile robot from 2d lidar scans using full convolutional neural

networks for security in cluttered environments," *Frontiers in neurorobotics*, vol. 12, p. 85, 2019.

[103] L. Martínez-Villaseñor, H. Ponce, J. Brieva, E. Moya-Albor, J. Núñez-Martínez, and C. Peñafort-Asturiano, "Up-fall detection dataset: A multimodal approach," *Sensors*, vol. 19, no. 9, p. 1988, 2019.

[104] S. Moulik and S. Majumdar, "Fallsense: An automatic fall detection and alarm generation system in iot-enabled environment," *IEEE Sensors Journal*, vol. 19, no. 19, pp. 8452–8459, 2018.

[105] G. Mastorakis and D. Makris, "Fall detection system using kinect's infrared sensor," *Journal of Real-Time Image Processing*, vol. 9, no. 4, pp. 635–646, 2014.

[106] S. Jankowski, Z. Szymański, U. Dziomin, P. Mazurek, and J. Wagner, "Deep learning classifier for fall detection based on ir distance sensor data," in *2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, vol. 2. IEEE, 2015, pp. 723–727.

[107] Y. Karayaneva, S. Sharifzadeh, W. Li, Y. Jing, and B. Tan, "Unsupervised doppler radar based activity recognition for e-healthcare," *IEEE Access*, vol. 9, pp. 62 984–63 001, 2021.

[108] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[109] S. Han, C. Qubo, and H. Meng, "Parameter selection in svm with rbf kernel function," in *World Automation Congress 2012*. IEEE, 2012, pp. 1–4.

[110] Y. Kim and H. Ling, "Human activity classification based on micro-doppler signatures using a support vector machine," *IEEE transactions on geoscience and remote sensing*, vol. 47, no. 5, pp. 1328–1337, 2009.

[111] A. Fleury, M. Vacher, and N. Noury, "Svm-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results," *IEEE transactions on information technology in biomedicine*, vol. 14, no. 2, pp. 274–283, 2009.

[112] Z. He and L. Jin, "Activity recognition from acceleration data based on discrete consine transform and svm," in *2009 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, 2009, pp. 5041–5044.

[113] M. Batool, A. Jalal, and K. Kim, "Sensors technologies for human activity analysis based on svm optimized by pso algorithm," in *2019 International Conference on Applied and Engineering Mathematics (ICAEM)*. IEEE, 2019, pp. 145–150.

[114] A. Sadiq, S. G. Khawaja, M. U. Akram, N. S. Alghamdi, A. Khan, and A. Shaukat, "Machine learning and signal processing based analysis of semg signals for daily action classification," *IEEE Access*, vol. 10, pp. 40 506–40 516, 2022.

[115] T. Malisiewicz, A. Gupta, and A. A. Efros, "Ensemble of exemplar-svms for object detection and beyond," in *2011 International conference on computer vision*. IEEE, 2011, pp. 89–96.

[116] M. A. Rahman, S. T. Hasan, and M. A. Kader, "Computer vision based industrial and forest fire detection using support vector machine (svm)," in *2022 International Conference on Innovations in Science, Engineering and Technology (ICISET)*. IEEE, 2022, pp. 233–238.

[117] P. S. Singh and S. Karthikeyan, "Salient object detection in hyperspectral images using deep background reconstruction based anomaly detection," *Remote Sensing Letters*, vol. 13, no. 2, pp. 184–195, 2022.

[118] N. Haider, "A review on object detection since 2005." *International Journal of Advanced Research in Computer Science*, vol. 13, no. 2, 2022.

[119] K.-M. Schneider, "A comparison of event models for naive bayes anti-spam e-mail filtering," in *10th Conference of the European Chapter of the Association for Computational Linguistics*, 2003.

[120] R. Swinburne, "Bayes' theorem," *Revue Philosophique de la France Et de l*, vol. 194, no. 2, 2004.

[121] I. Rish *et al.*, "An empirical study of the naive bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22, 2001, pp. 41–46.

[122] S. Taheri and M. Mammadov, "Learning the naive bayes classifier with optimization models," *International Journal of Applied Mathematics and Computer Science*, vol. 23, no. 4, pp. 787–795, 2013.

[123] Z. Feng, L. Mo, and M. Li, "A random forest-based ensemble method for activity recognition," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 5074–5077.

[124] A. U. Weerasuriya, X. Zhang, B. Lu, K. T. Tse, and C. Liu, "A gaussian process-based emulator for modeling pedestrian-level wind field," *Building and Environment*, vol. 188, p. 107500, 2021.

[125] S. Masarat, S. Sharifian, and H. Taheri, "Modified parallel random forest for intrusion detection systems," *The Journal of Supercomputing*, vol. 72, no. 6, pp. 2235–2258, 2016.

[126] P. Vepakomma, D. De, S. K. Das, and S. Bhansali, "A-wristocracy: Deep learning on wrist-worn sensing for recognition of user complex activities," in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2015, pp. 1–6.

[127] K. H. Walse, R. V. Dharaskar, and V. M. Thakare, "Pca based optimal ann classifiers for human activity recognition using mobile sensors data," in *Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 1*. Springer, 2016, pp. 429–436.

[128] N. Y. Hammerla, S. Halloran, and T. Plötz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," *arXiv preprint arXiv:1604.08880*, 2016.

[129] Y. Bengio, "Deep learning of representations: Looking forward," in *International conference on statistical language and speech processing*. Springer, 2013, pp. 1–37.

[130] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[131] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang, "Convolutional neural networks for human activity recognition using mobile sensors," in *6th international conference on mobile computing, applications and services*. IEEE, 2014, pp. 197–205.

[132] J. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition," in *Twenty-fourth international joint conference on artificial intelligence*, 2015.

[133] Y. Chen and Y. Xue, "A deep learning approach to human activity recognition based on single accelerometer," in *2015 IEEE international conference on systems, man, and cybernetics*. IEEE, 2015, pp. 1488–1492.

[134] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, "Deep convolutional neural networks and learning ecg features for screening paroxysmal atrial fibrillation patients," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 12, pp. 2095–2104, 2018.

[135] A. Sathyanarayana, S. Joty, L. Fernandez-Luque, F. Ofli, J. Srivastava, A. Elmagarmid, S. Taheri, and T. Arora, "Impact of physical activity on sleep: A deep learning based exploration," *arXiv preprint arXiv:1607.07034*, 2016.

[136] W. Jiang and Z. Yin, "Human activity recognition using wearable sensors by deep convolutional neural networks," in *Proceedings of the 23rd ACM international conference on Multimedia*, 2015, pp. 1307–1310.

[137] S. Ha, J.-M. Yun, and S. Choi, "Multi-modal convolutional neural networks for activity recognition," in *2015 IEEE International conference on systems, man, and cybernetics*. IEEE, 2015, pp. 3017–3022.

[138] M. S. Singh, V. Pondenkandath, B. Zhou, P. Lukowicz, and M. Liwickit, "Transforming sensor data to the image domain for deep learning—an application to footstep detection," in *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017, pp. 2665–2672.

[139] X. Li, Y. Zhang, I. Marsic, A. Sarcevic, and R. S. Burd, "Deep learning for rfid-based activity recognition," in *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*, 2016, pp. 164–175.

[140] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, "Deep learning for human activity recognition: A resource efficient implementation on low-power devices," in *2016 IEEE 13th international conference on wearable and implantable body sensor networks (BSN)*. IEEE, 2016, pp. 71–76.

[141] Y. Kim and B. Toomajian, "Hand gesture recognition using micro-doppler signatures with convolutional neural network," *IEEE Access*, vol. 4, pp. 7125–7130, 2016.

[142] T. Zebin, P. J. Scully, and K. B. Ozanyan, "Human activity recognition with inertial sensors using a deep learning approach," in *2016 IEEE sensors*. IEEE, 2016, pp. 1–3.

[143] S. Ha and S. Choi, "Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors," in *2016 international joint conference on neural networks (IJCNN)*. IEEE, 2016, pp. 381–388.

[144] M. Tschannen, O. Bachem, and M. Lucic, "Recent advances in autoencoder-based representation learning," *arXiv preprint arXiv:1812.05069*, 2018.

[145] B. Almaslukh, J. AlMuhtadi, and A. Artoli, "An effective deep autoencoder approach for online smartphone-based human activity recognition," *Int. J. Comput. Sci. Netw. Secur*, vol. 17, no. 4, pp. 160–165, 2017.

[146] A. Wang, G. Chen, C. Shang, M. Zhang, and L. Liu, "Human activity recognition in a smart home environment with stacked denoising autoencoders," in *International conference on web-age information management*. Springer, 2016, pp. 29–40.

[147] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[148] Y. Li, D. Shi, B. Ding, and D. Liu, "Unsupervised feature learning for human activity recognition using smartphone sensors," in *Mining intelligence and knowledge exploration*. Springer, 2014, pp. 99–107.

[149] H. Larochelle, M. Mandel, R. Pascanu, and Y. Bengio, "Learning algorithms for the classification restricted boltzmann machine," *The Journal of Machine Learning Research*, vol. 13, pp. 643–669, 2012.

[150] N. Y. Hammerla, J. Fisher, P. Andras, L. Rochester, R. Walker, and T. Plötz, "Pd disease state assessment in naturalistic environments using deep learning," in *Twenty-Ninth AAAI conference on artificial intelligence*, 2015.

[151] N. D. Lane, P. Georgiev, and L. Qendro, "Deepear: robust smartphone audio sensing in unconstrained acoustic environments using deep learning," in *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, 2015, pp. 283–294.

[152] T. Plötz, N. Y. Hammerla, and P. L. Olivier, "Feature learning for activity recognition in ubiquitous computing," in *Twenty-second international joint conference on artificial intelligence*, 2011.

[153] V. Radu, N. D. Lane, S. Bhattacharya, C. Mascolo, M. K. Marina, and F. Kawsar, "Towards multimodal deep learning for activity recognition on mobile devices," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, 2016, pp. 185–188.

[154] H. Fang and C. Hu, "Recognizing human activity in smart home using deep learning algorithm," in *Proceedings of the 33rd chinese control conference*. IEEE, 2014, pp. 4716–4720.

[155] L. Zhang, X. Wu, and D. Luo, "Real-time activity recognition on smartphones using deep neural networks," in *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*. IEEE, 2015, pp. 1236–1242.

[156] M. Edel and E. Köppe, "Binarized-blstm-rnn based human activity recognition," in *2016 International conference on indoor positioning and indoor navigation (IPIN)*. IEEE, 2016, pp. 1–7.

[157] Y. Guan and T. Plötz, "Ensembles of deep lstm learners for activity recognition using wearables," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 2, pp. 1–28, 2017.

[158] M. Inoue, S. Inoue, and T. Nishida, "Deep recurrent neural network for mobile human activity recognition with high throughput," *Artificial Life and Robotics*, vol. 23, no. 2, pp. 173–185, 2018.

[159] V. Miškovic *et al.*, "Machine learning of hybrid classification models for decision support," *Sinteza 2014-Impact of the Internet on Business Activities in Serbia and Worldwide*, pp. 318–323, 2014.

[160] F. J. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.

[161] X. Zhang, F. Chen, and R. Huang, "A combination of rnn and cnn for attention-based relation classification," *Procedia computer science*, vol. 131, pp. 911–917, 2018.

[162] F. Khozeimeh, D. Sharifrazi, N. H. Izadi, J. H. Joloudari, A. Shoeibi, R. Alizadehsani, J. M. Gorriz, S. Hussain, Z. A. Sani, H. Moosaei *et al.*, "Combining a convolutional neural network with autoencoders to predict the survival chance of covid-19 patients," *Scientific Reports*, vol. 11, no. 1, pp. 1–18, 2021.

[163] Y. Zheng, Q. Liu, E. Chen, Y. Ge, and J. L. Zhao, "Exploiting multi-channels deep convolutional neural networks for multivariate time series classification," *Frontiers of Computer Science*, vol. 10, no. 1, pp. 96–112, 2016.

[164] C. Liu, L. Zhang, Z. Liu, K. Liu, X. Li, and Y. Liu, "Lasagna: Towards deep hierarchical understanding and searching over mobile sensing data," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, 2016, pp. 334–347.

[165] C. Taramasco, T. Rodenas, F. Martinez, P. Fuentes, R. Munoz, R. Olivares, V. H. C. De Albuquerque, and J. Demongeot, "A novel monitoring system for fall detection in older people," *IEEE Access*, vol. 6, pp. 43 563–43 574, 2018.

[166] J. M. Quero, M. Burns, M. A. Razzaq, C. Nugent, and M. Espinilla, "Detection of falls from non-invasive thermal vision sensors using convolutional neural networks," *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 2, no. 19, 2018. [Online]. Available: https://www.mdpi.com/2504-3900/2/19/1236

[167] T. Li, B. Yang, and T. Zhang, "Human action recognition based on state detection in low-resolution infrared video," in *Proc. IEEE 16th Conf. on Ind. Electronics and Applications (ICIEA)*, 2021, pp. 1667–1672.

[168] S. Tateno, F. Meng, R. Qian, and Y. Hachiya, "Privacy-preserved fall detection method with three-dimensional convolutional neural network using low-resolution infrared array sensor," *Sensors*, vol. 20, no. 20, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/20/5957

[169] C. Perra, A. Kumar, M. Losito, P. Pirino, M. Moradpour, and G. Gatto, "Monitoring indoor people presence in buildings using low-cost infrared sensor array in doorways," *Sensors*, vol. 21, no. 12, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/12/4062

[170] J. Tanaka, H. Imamoto, T. Seki, and M. Oba, "Low power wireless human detector utilizing thermopile infrared array sensor," in *SENSORS, 2014 IEEE*, 2014, pp. 462–465.

[171] J. Tanaka, M. Shiozaki, F. Aita, T. Seki, and M. Oba, "Thermopile infrared array sensor for human detector application," in *2014 IEEE 27th International Conference on Micro Electro Mechanical Systems (MEMS)*, 2014, pp. 1213–1216.

[172] A. A. Trofimova, A. Masciadri, F. Veronese, and F. Salice, "Indoor human detection based on thermal array sensor data and adaptive background estimation," *Journal of Computer and Communications*, vol. 5, no. 4, pp. 16–28, 2017.

[173] S. Munir, S. Mohammadmoradi, O. Gnawali, and C. P. Shelton, "Measuring people-flow through doorways using easy-to-install ir array sensors," Mar. 16 2021, uS Patent 10,948,354.

[174] M. Burns, P. Morrow, C. Nugent, and S. McClean, "Fusing thermopile infrared sensor data for single component activity recognition within a smart environment," *Journal of Sensor and Actuator Networks*, vol. 8, no. 1, p. 10, 2019.

[175] A. Hayashida, V. Moshnyaga, and K. Hashimoto, "The use of thermal ir array sensor for indoor fall detection," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2017, pp. 594–599.

[176] T. Liu, X. Guo, and G. Wang, "Elderly-falling detection using distributed direction-sensitive pyroelectric infrared sensor arrays," *Multidimensional Systems and Signal Processing*, vol. 23, no. 4, pp. 451–467, 2012.

[177] V. Pavlov, H. Ruser, and M. Horn, "Feature extraction from an infrared sensor array for localization and surface recognition of moving cylindrical objects," in *2007 IEEE Instrumentation & Measurement Technology Conference IMTC 2007*, 2007, pp. 1–6.

[178] A. Sixsmith and N. Johnson, "A smart sensor to detect the falls of the elderly," *IEEE Pervasive computing*, vol. 3, no. 2, pp. 42–47, 2004.

[179] H. M. Ng, "Poster abstract: Human localization and activity detection using thermopile sensors," in *2013 ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2013, pp. 337–338.

[180] Y. Ogawa and K. Naito, "Fall detection scheme based on temperature distribution with ir array sensor," in *2020 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2020, pp. 1–5.

[181] Y. Watanabe, S. Kurihara, and T. Sugawara, "Sensor network topology estimation using time-series data from infrared human presence sensors," in *SENSORS, 2010 IEEE*. IEEE, 2010, pp. 664–667.

[182] B. Song, H. Choi, and H. S. Lee, "Surveillance tracking system using passive infrared motion sensors in wireless sensor network," in *2008 International Conference on Information Networking*. IEEE, 2008, pp. 1–5.

[183] M. Samara, "Literature review of sensor fusion technology: For improved occupancy information in indoor spaces," 2017.

[184] D. Wang and J. Lee, "Convolution-based design for real-time pose recognition and character animation generation," *Wireless Communications and Mobile Computing*, vol. 2022, 2022.

[185] H.-W. Tzeng, M.-Y. Chen, and J.-Y. Chen, "Design of fall detection system with floor pressure and infrared image," in *2010 International Conference on System Science and Engineering*. IEEE, 2010, pp. 131–135.

[186] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2625–2634.

[187] S. Berlemont, G. Lefebvre, S. Duffner, and C. Garcia, "Class-balanced siamese neural networks," *Neurocomputing*, vol. 273, pp. 47–56, 2018.

[188] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. European conference on computer vision*, 2016, pp. 391–407.

[189] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE conf. on computer vision and pattern Recognit.*, 2018, pp. 9446–9454.

[190] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[191] A. E. Ilesanmi and T. O. Ilesanmi, "Methods for image denoising using convolutional neural network: a review," *Complex & Intelligent Systems*, vol. 7, no. 5, pp. 2179–2198, 2021.

[192] L. Fan, F. Zhang, H. Fan, and C. Zhang, "Brief review of image denoising techniques," *Visual Computing for Industry, Biomedicine, and Art*, vol. 2, no. 1, pp. 1–12, 2019.

[193] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. on Medical image computing and computer-assisted intervention*, 2015, pp. 234–241.

[194] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE conf. on computer vision and pattern recognition*, 2016, pp. 770–778.

[195] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[196] Q. Huang, C. Hsieh, J. Hsieh, and C. Liu, "Memory-efficient ai algorithm for infant sleeping death syndrome detection in smart buildings," *AI*, vol. 2, no. 4, pp. 705–719, 2021.

[197] W. Nogami, T. Ikegami, R. Takano, T. Kudoh *et al.*, "Optimizing weight value quantization for cnn inference," in *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, pp. 1–8.

[198] Y. Gong, L. Liu, M. Yang, and L. Bourdev, "Compressing deep convolutional networks using vector quantization," *arXiv preprint arXiv:1412.6115*, 2014.

[199] M. Yu, Z. Lin, K. Narra, S. Li, Y. Li, N. S. Kim, A. Schwing, M. Annavaram, and S. Avestimehr, "Gradiveq: Vector quantization for bandwidth-efficient gradient aggregation in distributed cnn training," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[200] Q. Huang, "Weight-quantized squeezenet for resource-constrained robot vacuums for indoor obstacle classification," *AI*, vol. 3, no. 1, pp. 180–193, 2022.

[201] B. Jacob, S. Kligys, B. Chen, M. Zhu, M. Tang, A. Howard, H. Adam, and D. Kalenichenko, "Quantization and training of neural networks for efficient integer-arithmetic-only inference," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2704–2713.

[202] E. J. Kirkland, "Bilinear interpolation," in *Advanced Computing in Electron Microscopy*. Springer, 2010, pp. 261–263.

Appendix A

Publication List

A.1 Journals

[1] K. A. Muthukumar, M. Bouazizi, and T. Ohtsuki, "An Infrared Array Sensor-Based Approach for Activity Detection, Combining Low-Cost Technology with Advanced Deep Learning Techniques," *Sensors*, 2022; 22(10):3898

[2] K. A. Muthukumar, M. Bouazizi, and T. Ohtsuki, "A Novel Hybrid Deep Learning Model for Activity Detection Using Wide-Angle Low-Resolution Infrared Array Sensor," *IEEE Access*, vol. 9, pp. 82563–82576, 2021.

A.2 Conferences Proceedings (without peer-review)

[1] K. A. Muthukumar, M. Bouazizi, and T. Ohtsuki, "Comparative Study of Activity Detection System Using Wide Angle Low-Resolution Infrared Array Sensor at Different Positions," *IEICE General Conf.*, BS–4–2, Mar. 2021.

[2] K. A. Muthukumar, M. Bouazizi, and T. Ohtsuki, "Activity Detection Using Wide Angle Low-Resolution Infrared Array Sensors," *IEICE Society Conf.*, BS–8–1, Sep. 2020.

A.3 Technical Reports

[1] K. A. Muthukumar, M. Bouazizi, and T. Ohtsuki, "Detection of human activity based on hybrid deep learning model using a low-resolution infrared array sensor," *IEICE Tech. Rep.*, vol. 120, no. 261, SeMI 2020–39, pp. 99–104, Nov. 2020.

A.4 Awards

[1] Sep. 2020 ICM English session award (IEICE ICM Section)