Enhancing Human Activity Recognition Through IoT Sensor Data Analytics: A Deep Learning Approach

Final Year Project Report

Presented

by

Sabir Jan
CIIT/FA20-BCE-075/ISB

Maheen Arshad
CIIT/ FA21-BCE-035/ISB

M.Naeem Farooq
CIIT/ FA21-BCE-052/ISB

In Partial Fulfillment

of the Requirement for the Degree of

Bachelors of Science in Computer Engineering

DEPARTMENT OF COMPUTER ENGINEERING

COMSATS UNIVERSITY ISLAMABAD

July 2025

Declaration

We hereby declare that this project neither as a whole nor as a part has been copied out from any source. It is further declared that we have developed this project and the report accompanied entirely on the basis of our personal efforts made under the sincere guidance of our supervisor. No portion of the work presented in this report has been submitted in the support of any other degree or qualification of this or any other University or Institute of learning, if found we shall stand responsible.

Signature: _______________
Sabir Jan

Signature: _______________
Maheen Arshad

Signature: _______________
M.Naeem Farooq

COMSATS UNIVERSITY ISLAMABAD
July 2025

Enhancing Human Activity Recognition Through IoT Sensor Data Analytics: A Deep Learning Approach

An Undergraduate Final Year Project Report submitted to the
Department of
COMPUTER ENGINEERING

As a Partial Fulfillment for the award of Degree

Bachelor of Science in Computer Engineering

by

Name	Registration Number
Sabir Jan	CIIT/FA20-BCE-075/ISB
Maheen Arshad	CIIT/FA21-BCE-035/ISB
M.Naeem Farooq	CIIT/FA21-BCE-052/ISB

Supervised by

Dr. Muhammad Dilshad Sabir
Assistant Professor
Department of Computer Engineering
CUI, Islamabad

COMSATS UNIVERSITY ISLAMABAD

July 2025

Final Approval

This Project Titled

Enhancing Human Activity Recognition through IoT sensor Data Analytics: A Deep Learning Approach

Submitted for the Degree of

Bachelor of Science in Computer Engineering

By

Name	**Registration Number**
Sabir Jan	CIIT/FA20-BCE-075/ISB
Maheen Arshad	CIIT/FA21-BCE-035/ISB
M. Naeem Farooq	CIIT/FA21-BCE-052/ISB

has been approved for

COMSATS UNIVERSITY ISLAMABAD

____

Supervisor
Dr. Muhammad Dilshad Sabir
Assistant Professor

____

Co-Supervisor
Dr. Samera Batool
Assistant Professor

____

Internal Examiner-1
Dr. Mahmood Qureshi Hashmi
Associate Professor

____

Internal Examiner-2
Dr. Guftaar Ahmad
Associate Professor

____

External Examiner
Name
Designation

____

Head
Department of Computer Engineering

Dedication

We would like to dedicate our work to Allah Almighty who has provided us the strength and courage to complete our work. Without his strength and courage this work would not have been possible.

We also extend our heartfelt appreciation to our esteemed supervisor, Dr. Muhammad Dilshad Sabir, and our co-supervisor, Dr. Samera Batool whose enthusiastic guidance and direction were essential in the completion of this project.

Lastly, we would like to dedicate our work to our parents, whose love and sacrifices have made our educational journey possible. Their belief in us has been our motivation to reach this point. This work reflects their endless support and encouragement, and we are forever grateful for everything they have done for us.

Acknowledgements

We extend our heartfelt thanks to our supervisors, Dr. Muhammad Dilshad Sabir and Dr. Samera Batool, for their invaluable direction, guidance, and mentorship throughout this project. Their expertise, insightful evaluations, and continuous encouragement have significantly enhanced the alignment of our work with its objectives and improved its overall quality.

We also express our sincere gratitude to our families for their unwavering support, patience, and most importantly, their deep understanding even during times when our ambitions seemed like mere dreams. Their faith in our efforts has been a constant source of motivation, driving us to see this project through to completion. They have been the solid foundation and a source of strength behind our success. Without them, navigating the challenges and celebrating the triumphs of this academic journey would not have been possible.

We are equally thankful to COMSATS University Islamabad for providing a nurturing environment for learning and research. The access to resources, facilities, and a collaborative atmosphere played a crucial role in the development and execution of this project.

Lastly, we extend our appreciation to everyone who contributed, directly or indirectly, to the success of this project. Your support has been truly meaningful.

Table of Contents

List of Acronyms

FER Facial Expression Recognition

HAR Human Activity Recognition

ER Emotion Recognition

CSI Channel State Information

CNNs Convolutional Neural Networks

EEG Electroencephalography

OFDM Orthogonal Frequency-Division Multiplexing

RSSI Received Signal Strength Indicator

IoT Internet of Things

LSTM Short-Term Memory

RNNs Recurrent Neural Networks

SVM Support Vector Machines

RF Random Forests

ANNs Artificial Neural Networks

TX Transmitter

RX Receiver

AP Access Point mode

List of Figures

List of Tables

Enhancing Human Activity Recognition Through IoT Sensor Data Analytics: A Deep Learning Approach

Final Year Project Report

Presented

by

Sabir Jan
CIIT/FA20-BCE-075/ISB

Maheen Arshad
CIIT/ FA21-BCE-035/ISB

M.Naeem Farooq
CIIT/ FA21-BCE-052/ISB

In Partial Fulfillment

of the Requirement for the Degree of

Bachelors of Science in Computer Engineering

DEPARTMENT OF COMPUTER ENGINEERING

COMSATS UNIVERSITY ISLAMABAD

July 2025

Abstract

Facial Expression Recognition (FER) is an important field of human-computer interaction that finds applications in healthcare, security, biometric authentication, and human-robot interaction. The thesis proposes a new FER method by utilizing Wi-Fi-based Channel State Information (CSI) for the detection and classification of facial expressions without explicit visual observation. The system design consists of a Wi-Fi installation which sends and receives CSI signals, which are affected by minor facial muscle movements. These signals are collected through data acquisition and preprocessing, such as feature extraction, and labeling of five basic expressions: Neutral, Happy, Sad, Fear, and Surprise. Thereafter, machine learning models are trained and tested using this preprocessed data to label facial expressions with high accuracy. The classification results are further used in different application domains, highlighting the potential of the system in privacy-preserving and non-intrusive FER. This work helps advance intelligent systems capable of interpreting human emotions without excessive dependency on camera-based surveillance, thus improving user experience and system intelligence in several real-world applications.

Chapter 1. Introduction

Human Activity Recognition (HAR) is all about using sensor data to automatically figure out what people are doing-like walking, sitting, or exercising. This technology has become important for making smart systems that can respond to what people need in real time. For example, HAR is used in healthcare to detect falls in elderly people, in security systems, in smart homes, and even for fitness and entertainment apps[1].

Emotion Recognition (ER) tries to understand how people are feeling by looking at things like their behavior, body signals, or facial expressions. Most ER systems today use cameras to read facial expressions, but this can make people feel uncomfortable and worried about their privacy and data security.

Recently, new wireless technologies have made it possible to use radio signals-like Wi-Fi-to pick up on small changes in the environment caused by people's movements or expressions. By analyzing how Wi-Fi signals change CSI, we can detect activities and emotions without needing cameras or wearable devices. This means we can monitor what's happening in a room in a way that's private, doesn't need people to wear anything special, and fits easily into everyday life [2].

1.1 Purpose And Overview

In recent years, the fields of Human Activity Recognition (HAR) and Emotion Recognition (ER) have become very important due to their wide range of applications in healthcare monitoring, security purposes, smart homes, and human-computer interaction systems. In the past, these fields have relied on camera setups or wearable sensors for data collection. While these are effective to a certain extent, these methods have critical challenges, including invasions of user privacy, limitations in various environmental conditions, and the necessity for continuous user compliance. The goal of this project is to make things simpler for the users. For this purpose, we will use Wi-Fi signals and data from smart sensors in a way that protects people's privacy. Specifically, we focus on detecting five key human emotions — happy, sad, fear, surprised, and neutral — solely by analyzing variations in Wi-Fi signals.

Understanding someone's emotions gives us more context than just knowing what they are doing. For example, knowing someone is sitting is good, but knowing they are sitting and sad gives us even better information. This can be very useful in things like checking on mental health or giving people personalized help. By including facial emotion detection, this project tries to create a better system that can understand people in a more complete way. ESP32 chips are a good choice because they are not too expensive and work well for gathering and processing the data needed to recognize emotions.

This system uses two small Wi-Fi devices (ESP32) configured as a transmitter-receiver pair to collect CSI data. Next the process is followed by comprehensive data preprocessing, feature extraction, and model training. Through this work, we aim to develop a practical, scalable, and privacy-respecting solution that can function effectively in dynamic real-world environments.

Figure 1: Facial Expressions

1.1.1 Challenges in Human Activity and Emotion Recognition

Despite the fast progress in Human Activity Recognition (HAR) and Emotion Recognition (ER), there are still some challenges that make it difficult for these technologies to be used everywhere.

Privacy and Ethics:

Systems that use cameras often capture sensitive images and videos, which can invade people's privacy-especially in private spaces like homes or hospitals. This makes many people uncomfortable and raises serious ethical questions [3].

Sensitive to the Environment:

Wi-Fi signals can be affected by things like walls, furniture, and other electronic devices. These "noises" can mess up the data, making it harder for the system to correctly recognize activities or emotions.

Reliance on Devices:

Wearable systems expect people to always wear their devices properly. This can be tough for long-term use, especially for groups like the elderly, who may forget or find it uncomfortable to keep wearing these gadgets.

Detecting Small Emotional Changes:

Facial expressions often involve tiny, quick movements. These are much harder for Wi-Fi-based systems to pick up compared to bigger body movements, making emotion recognition more challenging.

Working in Real Life:

Many current Wi-Fi sensing systems work well in labs but struggle in real-world places where there are lots of moving objects, changing conditions, and other sources of interference.
Our project aims to tackle these problems by building a strong data processing system and using machine learning that's specially designed to pick up on the small changes in Wi-Fi data linked to both activities and emotions. This way, we hope to create a solution that is practical, accurate, and respects people's privacy-even in busy, everyday environments.

1.2 Problem Statement

Human Activity Recognition (HAR) and Emotion Recognition (ER) technologies hold great promises for transforming areas like personalized healthcare, smart living spaces, and systems that put people at the center of technology. Traditionally, most of these systems have relied on either camera-based setup, which can make people uneasy about their privacy, or wearable sensors, which require users to consistently wear and interact with devices. Over time, this can lead to discomfort and issues with people not using the devices as intended [2].
Recently, Wi-Fi sensing-especially through analyzing Channel State Information (CSI)-has emerged as an exciting alternative. This approach can pick up on subtle changes in the environment caused by human movement or emotion, without the need for cameras or wearables. However, using Wi-Fi signals for this purpose isn't without its own challenges. The accuracy and reliability of these systems can be affected by things like interference from other signals, background noise in the environment, and differences in hardware. These problems become even more pronounced in real-world settings, where conditions are far less controlled than in a lab.
Because of these challenges, there is a real need for a solution that can accurately and reliably recognize human activities and emotions without invading privacy or requiring people to wear special devices-and that can still work well even when the environment is unpredictable.
This thesis aims to fill that gap by introducing a method that uses CSI data from Wi-Fi signals, combined with machine learning, to improve the detection and classification of both activities and emotions. The goal is to achieve high levels of accuracy and reliability, while making sure the system is non-intrusive and comfortable for users, and protects their privacy, even in the messy, dynamic conditions of everyday life.

1.3 Proposed Solution

To rectify the deficits of existing systems, we present a new approach that leverages Wi-Fi Channel State Information (CSI) and IoT sensor readings to enhance how we identify human activities and emotions. The proposed solution is practical, privacy-friendly, and functional in real-world settings, and it unfolds over several significant phases. First, a dataset of about 2,500 samples would be collected, spread over five basic emotions-happy, sad, fear, surprised, and neutral-and a few physical activities such as sitting, walking, and running. Data collection is achieved by the help of two ESP32 devices: one sending Wi-Fi signals and the other receiving waves, to sense or capture small changes from the movements and expressions.

Since the raw CSI data is very much likely to be dirty due to background noise and other environmental interferences, purification of the data is the next step. We will be applying sufficiently strong preprocessing techniques-such as filtering, denoising, and normalization-to ensure that the signals are as clear and accurate as possible56. Next is the feature extraction. There, we will extract the most important pieces of information from the cleaned data- those that reflect the variations of the facial expressions. We will consider statistical features as well as patterns in both time and frequency domains, thus ensuring the capture of all relevant details [2].

Then comes the machine learning part, in which we'll train and test various algorithms from Super Learner to Random Forests and Convolutional Neural Networks (CNNs) to identify which works best when recognizing activities or emotions as derived from the data. [1] To ensure that the models are credible and do not just memorize the training data, different cross-validation techniques would be used to check on performance regarding novel unseen data Finally, we'll evaluate how well the system works by measuring accuracy, precision, recall, F1 score, and by looking at confusion matrices.

These metrics will give us a clear picture of the system's strengths and any areas that need improvement. By adhering to this systematic approach, our system of proposal seeks to provide high-precision recognition of activities and emotions alike, and be scalable, user-friendly, and privacy-respecting-yet another excellent match for real-world applications [3].

1.3 Key Objectives

The main goals behind developing this project are centered on making human activity and emotion recognition easier, more private, and practical for everyday use. First, we want to build a system that can recognize what people are doing and how they're feeling-without needing cameras or wearable devices. This way, people's privacy is protected, and there's no need for anyone to wear extra gadgets. To make this possible, we'll collect a high-quality dataset using Wi-Fi Channel State Information (CSI) signals. This dataset will include around 2,500 samples, covering five key emotions (happy, sad, fear, surprised, and neutral) as well as common activities like sitting, walking, and running, all recorded in different controlled settings. The data will be gathered using two ESP32 devices-one sending out Wi-Fi signals and the other receiving them-to capture the changes caused by different movements and expressions. Once we have the data, we'll focus on cleaning it up. Wi-Fi signals can easily get mixed up with noise and interference from the environment, so we'll use strong preprocessing methods like filtering and normalization to make sure the data is as clear and accurate as possible. Next, we'll extract the most important features from the cleaned data-those that best represent the differences between various emotions and activities. We'll look at different types of features, including those based on statistics, time, and frequency, to make sure we capture all the useful details.

We'll then train and test several machine learning models, such as Support Vector Machines, Random Forests, and Convolutional Neural Networks, to see which one works best for this kind of data. We'll use techniques like cross-validation to make sure our models are reliable and can handle new situations, not just the examples they were trained in. To measure how well our system works, we'll use a range of performance metrics, including accuracy, precision, recall, F1 score, and confusion matrices. This will give us a clear, well-rounded view of how effective our approach is compared to traditional camera-based or wearable systems, as well as other Wi-Fi sensing methods [5]. Finally, we'll explore how this system can be used in real-world areas like healthcare, smart homes, and security, where privacy and ease of use are especially important. By following this approach, we hope to create a solution that is not only accurate and reliable, but also comfortable and practical for people in their daily lives.

1.4 Benefits of the Project

Enhanced User Privacy:

The biggest advantage of this system is that it doesn't rely on cameras, which means it won't invade people's personal spaces or capture sensitive images. This makes it much more ethical and acceptable for use in private areas like homes, hospitals, or offices, where people want to feel secure.

Non-Intrusive and Comfortable:

Unlike wearable devices that people must remember to wear or interact with, this system works invisibly in the background. Users can go about their daily routines naturally-no need to adjust behavior, wear gadgets, or worry about forgetting a device.

Affordable and Easy to Scale:

We use low-cost ESP32 devices (common, budget-friendly hardware) and existing Wi-Fi networks, which keep costs down. This makes it practical to deploy the system widely, whether in homes, hospitals, schools, or public spaces, without needing expensive infrastructure.

Handles Real-World Environments Well:

Even though Wi-Fi signals can get messy due to noise, interference, or changing conditions (like moving furniture or people), our system uses advanced techniques to clean up the data and focus on what matters. This helps it stay reliable even in unpredictable, everyday settings.

Useful Across Many Areas:

This technology isn't limited to one field. It can help monitor patients in healthcare without constant check-ins, detect unusual activity or emotional distress in security systems, or adjust smart home devices (like lights or thermostats) based on how people are feeling or moving.

Pushing Technology Forward:

By using Wi-Fi Channel State Information (CSI)-which is far more detailed than basic signal strength (RSS)-we're unlocking new ways to recognize activities and emotions with wireless signals. This could lead to smarter, more intuitive systems that work seamlessly in the background of our lives.

In short, this project isn't just about avoiding cameras or wearables-it's about creating a system that's practical, respectful, and adaptable enough to fit into the real world without compromising on performance or privacy.

1.5 Motivation

The motivation for this project comes from what people need in today's world: technology that helps us live better lives without making us feel watched or uncomfortable. As everything around, us gets more connected and digital, it's more important than ever to have systems that can look for our well-being but still respect our privacy. People want solutions that simply provide useful tasks quieting down in the background, augmenting us in a very harmonious way, which does not impede or makes us feel under constant observation. There is a sense of too much invasion of privacy. Cameras or wearable sensors usually create feelings of being watched, how our personal information might be used, and the perception of losing control over their own lives.

The COVID-19 pandemic made it even clearer how valuable it is to have contactless ways to monitor health, especially when people are isolated or can't be with others in person. What's exciting is that we already have Wi-Fi everywhere, and we can use it for much more than just internet access. We can develop smart systems that detect activity and emotion without any additional devices or complex setups simply by tapping existing Wi-Fi signals. Very much blended with today's lifestyle, technology makes things easier and comfortable for everyone.

It's about two things primarily: pushing so that technology can again help people in a more natural and unobtrusive way; and making certain that while new systems like these are being constructed, privacy and freedom are kept in mind. This will result in systems that are not simply intelligent but also worthy of trust and respect for users.

1.6 Broader Impact (UN SGDs)

1.6.1 Good Health and Well-being (Goal 3)

The system helps in monitoring emotional states and physical activities, which is useful for mental health care and patient support. It can also help detect emergencies like falls or unusual behavior, especially for elderly or sick individuals, making it easier to take quick action.

1.6.2 Industry, Innovation, and Infrastructure (Goal 9)

By using modern technologies like machine learning, wireless sensing, and ESP32 devices, this project promotes new and smart ways of building health and safety systems. These technologies are low-cost and can be improved and used in many different industries.

1.6.3 Sustainable Cities and Communities (Goal 11)

The system can be used in smart homes and cities to make them safer and more responsive to people's needs. It does not use cameras, so it protects people’s privacy while still being able to monitor their activities and emotions.

1.6.4 Reduced Inequalities (Goal 10)

The affordable nature of ESP32-based solutions makes this technology accessible to a broader population, helping reduce the technological divide. It supports inclusive monitoring systems that can be used in under-resourced communities or regions lacking access to advanced healthcare technology.

1.6.5 Peace, Justice, and Strong Institutions (Goal 16)

This project addresses ethical concerns around surveillance and privacy. By avoiding the use of cameras and providing transparent, secure data handling, it supports the development of trustworthy AI systems that respect user rights.

Chapter 2. Literature Review

The advancement of sensing technologies has enabled novel, non-invasive methods for detecting human activities and emotions. Traditional approaches rely heavily on visual or wearable sensors, but these techniques often come with privacy, intrusiveness, and reliability issues. Recently, Wi-Fi Channel State Information (CSI) has emerged as a promising solution, offering contactless sensing capabilities for applications in human activity recognition (HAR) and emotion recognition (ER). This literature review presents a comprehensive analysis of the recent research trends in this field, focusing on HAR, ER, the use of CSI, comparison with RSSI, sensor technologies, the role of wearable versus contactless systems, directional antennas, IoT-based systems, and limitations.

2.1 Human Activity Recognition (HAR)

Human activity recognition has traditionally been based on vision-based sensors or wearable data. Recent developments in wireless technologies, especially in Wi-Fi CSI, have provided new mechanisms for sensing motion through signal variation analysis. Li et al. introduced Wi-Motion, a system that uses CSI to detect different human movements robustly with outstanding accuracy [5].

Wi-Motion is an apparatus that employs Wi-Fi signals to identify various human movements. It identifies movements by examining Channel State Information (CSI), which indicates the variations in Wi-Fi signals as they travel through the environment. If a person moves, his or her body interferes with the signals, and Wi-Motion can detect these variations to determine unique movements. This allows one to detect movements without cameras or wearable devices, making the system more personal and convenient.

Likewise, Zhuravchak et al. proposed a deep neural network model based on Wi-Fi CSI data for HAR, showing that CSI can separate fine-grained activities, including walking, sitting, or falling, without the need for line-of-sight contact [18].

Moshiri et al. also made a major contribution to this field with several works. They showed, in [16] that CNNs learned from CSI signals could effectively recognize various indoor activities. Shahverdi et al. also verified that CNNs are suitable for CSI-based HAR and have potential in real-time applications [14].

2.2 Emotion Recognition (ER)

Emotion recognition is vital for enhancing human-computer interaction, mental health monitoring, and adaptive systems. Traditionally, ER has employed facial expressions, speech, EEG signals, and physiological responses. Pise et al. reviewed methods to recognize facial expressions in challenging situations, highlighting the limitations of visual-only systems in real-world conditions [1]. Wu et al. introduced an Edge-AI-driven facial expression recognition system optimized for mobile devices, showing its relevance for embedded applications [6].

In parallel, wireless signal-based emotion recognition is gaining traction. Zhao, Adib, and Katabi showcased how Wi-Fi signals could be analyzed to detect subtle body changes linked with emotions, using variations in wireless reflections to interpret mental states [10]. Gu et al. combined Wi-Fi and visual data for emotion recognition using facial gestures, creating a multimodal approach for improved robustness [3].

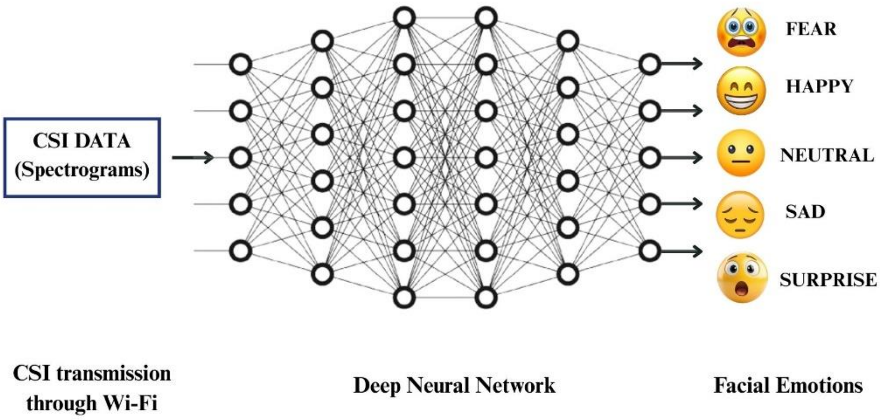

Figure 2: Facial Emotion Recognition system

2.3 Wi-Fi and Channel State Information (CSI)

Emotion recognition is crucial to improving human-computer interaction, mental health monitoring, and adaptive systems. Facial expressions, speech, EEG signals, and physiological responses have been traditionally used in ER. Pise et al. surveyed approaches to facial expression recognition in adverse conditions, emphasizing the shortcomings of visual-only systems in real-world environments [1]. Wu et al. proposed an Edge-AI-driven facial expression recognition system designed for mobile devices, demonstrating its applicability for embedded applications [6].

Concurrently, emotion recognition using wireless signals is also gaining prominence. Zhao, Adib, and Katabi demonstrated how Wi-Fi signals can be examined to identify subtle body movements associated with emotions, utilizing changes in wireless reflections to decode mental states [10]. Gu et al. integrated Wi-Fi and visual data for facial gesture-based emotion recognition, developing a multimodal framework for enhanced robustness [3].

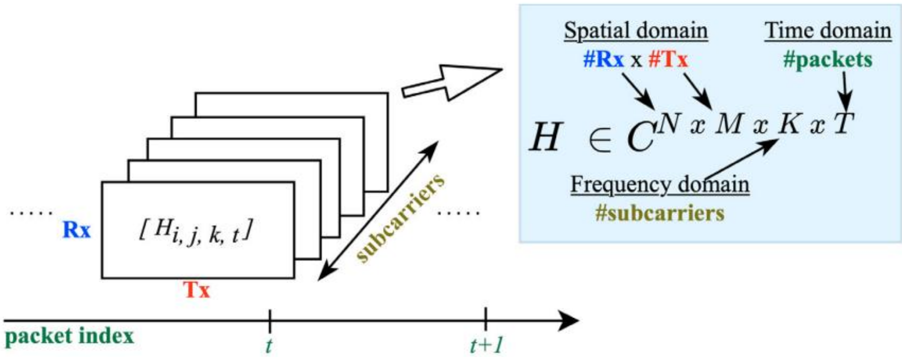

Figure 3: 4D CSI tensors is a time series of CSI matrices of MIMO-OFDM channels [18]

Hou et al. proposed a multimodal emotion recognition framework using both Wi-Fi CSI and visual information, showcasing CSI's capability in emotion detection when combined with other data modalities [4]. This system watches someone's body and faces understanding them. It uses tiny radar waves to track their heartbeat and breath without touching them. A regular camera records their facial expressions. To isolate just the individual, it eliminates background noise from the equation. Then, it distinguishes the breathing and heartbeat signals so that it can analyze them. Essentially, it couples radar body sensing with face monitoring to gain a better understanding of what someone's state is. Pham et al. created a CSI improvement framework based on machine learning to automatically segment and preprocess CSI data to improve recognition performance [13].

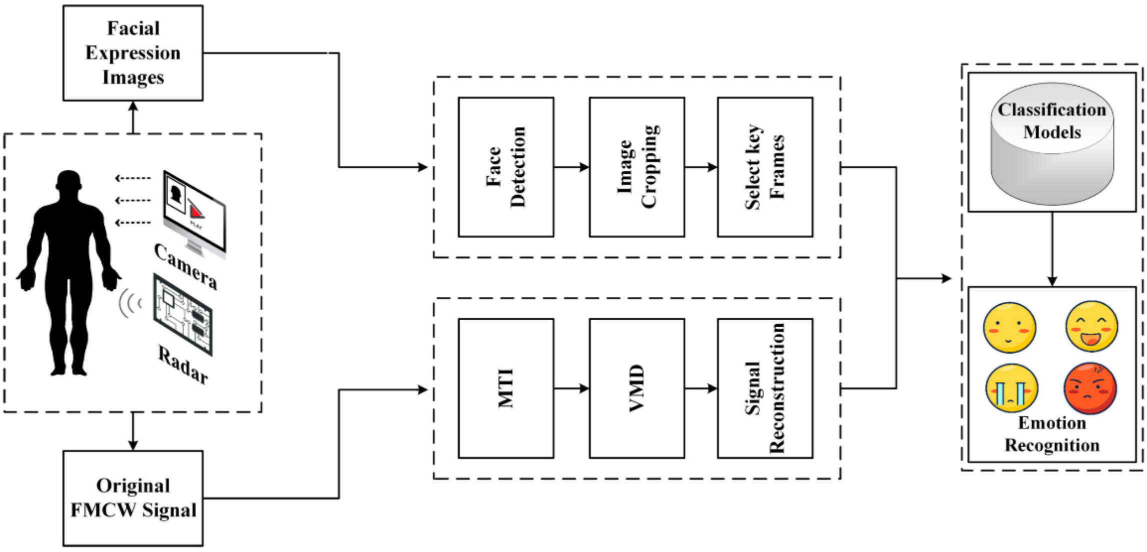

Figure 4: The overall frame diagram of the Emotion Recognition using Multimodal Information [19]

2.4 CSI vs RSSI

2.4.1 CSI

All CSI measurements are obtained through ESP32 microcontrollers with the assistance of the ESP32-CSI Toolkit [12]. Channel State Information (CSI) is a signal measurement in systems that send data via Orthogonal Frequency-Division Multiplexing (OFDM). It is a description of how the amplitude and phase of the signal vary as it propagates from a transmitter to a receiver through various frequency bands, referred to as subcarriers. In the frequency domain, the OFDM system can be represented by the equation:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \qquad \textbf{(1)}$$

where **x** is the original signal sent, **y** is the signal received, **n** is the noise, and **H** is a matrix that represents how the wireless channel has affected the signal — this is the CSI. CSI is recorded for each subcarrier, and each value is a complex number with a real part $\mathbf{H(i)_r}$ and an imaginary part $\mathbf{H(i)_{im}}$.

With that, denoting the subcarrier index as i, we can calculate **amplitude** from raw CSI,

$$A^{(i)} = \sqrt{(H_{im}^{(i)})^2 + (H_r^{(i)})^2} \qquad \textbf{(2)}$$

and phase as well,

$$\Phi^{i} = \mathrm{atan2}(H_{im}^{(i)}, H_r^{(i)}) \qquad \textbf{(3)}$$

This detailed information helps in understanding how the environment affects Wi-Fi signals.

2.4.2 RSSI

RSSI represents a coarse-grained measurement that provides a single value indicating the strength of received signals. It is measured by a packet index and offers a simplified view of the wireless channel. [20] The primary advantage of RSSI lies in its widespread availability and ease of implementation across various wireless devices. RSSI distance measurement generally uses the logarithmic distance path-loss model.

In contrast to traditional RSSI, CSI describes the small-scale multipath fading and therefore serves as a more detailed descriptor of the wireless channel [21]. Dzedzickis et al. presented a comparative overview of sensing technologies and concluded that CSI excels RSSI in detecting complicated activities and physiological signals [8]. The benefit of CSI is the capability to detect very subtle signal changes owing to minor body movements, making CSI well-suited for emotion and activity recognition.

2.5 Wearable Devices vs Contactless Devices

Wearable sensors like EEG headbands and smartwatches have also been extensively employed for HAR and ER. Tao et al. applied channel-wise and self-attention-based EEG-based emotion recognition, which emphasized the efficacy of physiological signals in emotional state analysis [2]. Wearable sensors, though, need habitual user compliance, e.g., remembering to wear and charge them on a daily basis, which is cumbersome. Moreover, they can be intrusive or bothersome to certain users, possibly interfering with daily operations or posing privacy issues. On the other hand, contactless systems utilizing audiovisual aids (such as microphones and cameras) or wireless sensing techniques (such as Wi-Fi signals) provide a less burdensome and more comfortable option for users. They don't require wear or contact, which is less irritating and more convenient for individuals to utilize in everyday life.

For instance, Komagal and Yogameena developed a facial expression detection system that utilizes a camera to monitor students within a classroom environment. This system is able to identify emotions without requiring the students to wear any device or touch a sensor physically [7]. This simplifies its use within environments such as schools where large numbers of individuals are involved, and it's not feasible for everyone to wear a device.

The shift towards contactless approaches—particularly those that are based on Wi-Fi signals and Channel State Information (CSI)—has numerous benefits. Firstly, it better safeguards user privacy as it doesn't capture personal images or sounds themselves. Secondly, it is simpler to implement and use in expansive spaces or smart spaces, like offices, residences, and public areas, because it doesn't require users to wear or carry any devices.

2.6 Long-Range Sensing with Directional Antennas

Directional antennas are important for increasing the range, precision, and concentration of Wi-Fi-based sensing systems. By focusing the signal in one direction, these antennas minimize signal dispersion and enhance the quality and strength of the received data. Strohmayer and Kampel performed thorough research on directional antenna configurations' long-range through-wall sensing in detecting and observing human activity by penetrating solid structures like walls and furniture [9]. Their results depict the potential to use directional antennas to facilitate powerful Human Activity Recognition (HAR) and Emotion Recognition (ER) with no need to use invasive strategies such as cameras or wearable units. Such sophisticated configurations best fit use in smart homes, health monitoring, and security surveillance applications where non-contact and privacy-friendly solutions are greatly in demand.

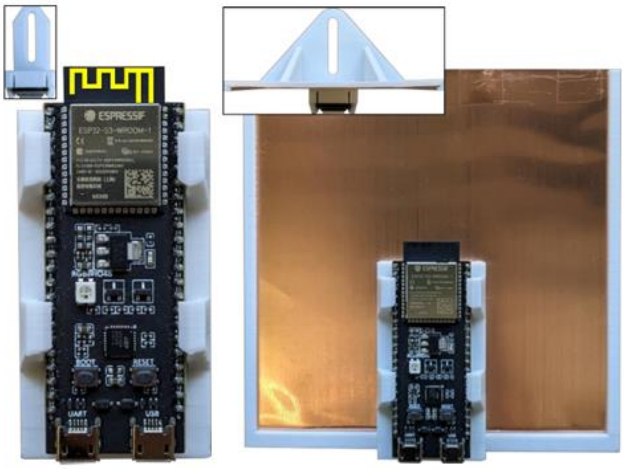

Figure 5: PIFA & PIFA with a plane reflector [10]

2.7 Lightweight and IoT-Compatible Hardware

HAR and ER systems must be lightweight and IoT-compatible to enable deployment on a large scale. Hernandez and Bulut presented an independent IoT-based system for mobility sensing and emphasized its simple deployment and power efficiency [12]. Tong et al. also presented human activity recognition on embedded platforms using low-cost hardware and demonstrated that sophisticated recognition is possible even on limited resources platforms [10].

El Zein et al. presented a lightweight CNN model coupled with data augmentation for HAR, which balances model accuracy and computational cost [16]. Contributions of this nature are essential in real-time smart environment applications.

2.7.1 Comparison among CSI Collecting Hardware

There is many hardware platforms from which CSI can be collected. Intel 5300 NIC and Atheros chipsets are usually used to tap CSI from Wi-Fi packets. ESP32 devices have been reworked for low-cost CSI sensing. Moshiri et al. used various such platforms in their research to analyze CSI-based HAR, and their comparison informed them of what configurations provide the optimal tradeoff between cost, accuracy, and scalability [16].

Pham et al.'s CSI enhancement framework also facilitates effective CSI extraction, preprocessing, and classification, particularly for real-time systems [13].

	USRP	Atheros	Intel 5300	ESP32
Cost	$8,400	$11	$11	$6
Size (cm)	26.7×21.8	2.98×2.82	3×2.68	5.5×2.8
Weight	1.6kg	>1kg (*)	>1kg (*)	10g

Table 1: Comparison between CSI collecting hardware [11]

(*) Including the computer.

2.8 Deep Learning Models for Emotion Recognition

Deep learning is now the pillar of current research in emotion recognition, especially with the increased access to heterogeneous sensing modalities like vision, Wi-Fi, and EEG signals. Convolutional Neural Networks (CNNs) are now among the most common deep learning structures for emotion and activity recognition due to their excellent feature extraction capacity. For example, in their research on WiFi CSI-based activity recognition, Shahverdi et al. employed CNNs to extract spatial features from signal data and showed their resilience across different environments [14]. Likewise, Moshiri et al. confirmed the effectiveness of CNNs for activity and emotion recognition by taking advantage of deep learning on Channel State Information (CSI) and obtained high classification accuracy using both regular and lightweight CNN models [16].

A number of works have built upon conventional CNNs by incorporating attention mechanisms to enhance model interpretability and accuracy. Tao et al. introduced a dual attention approach for EEG-based emotion recognition, including both channel-wise and self-attention modules to weigh features dynamically, and this greatly enhanced recognition accuracy between emotion classes [2]. This application of attention mechanisms points to a trend in the field where attention layers are added to neural networks to deal with high-dimensional data and enhance feature selection.

Multimodal methods that blend vision-based inputs with Wi-Fi signals have also become trendy. Hou et al. created a system that combines Wi-Fi and visual features through deep learning to obtain better emotion detection in challenging scenarios [4]. Gu et al. introduced WiFE, a system that incorporates Wi-Fi CSI and facial gesture recognition through computer vision and deep learning, demonstrating that hybrid models perform better than unimodal systems, especially in real-world environments with occlusions or low light conditions [3].

Some other models that have been promising are Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, especially for time-series emotion data like EEG or CSI signals. But their application has declined marginally in favor of CNNs as the latter train faster and are less complex. Lightweight CNNs have been highlighted in recent literature for edge device deployment. Wu et al. suggested an Edge-AI architecture integrating effective CNN models with mobile architecture for real-time facial expression recognition, which ensured that performance can be retained while lowering the computational cost [6]. It is especially suitable for embedded or IoT-based applications.

Classic machine learning models like Support Vector Machines (SVM) and Random Forests (RF) were also tried by Hameed et al., [22], but deep models like Artificial Neural Networks (ANNs) and EfficientNetV2 performed much better, particularly in interpreting intricate non-visual patterns of data from RF input. Strohmayer and Kampel [9] employed EfficientNetV2, which is a leading-edge CNN structure renowned for striking a balance between parameter efficiency and accuracy.

Moreover, automatic segmentation and data augmentation techniques have been used to improve model generalization. Pham et al. proposed a CSI signal improvement framework that integrates machine learning-based automatic segmentation, which significantly improves classification performance without the need for heavy manual preprocessing [13].

Among the most common models utilized in studies are CNNs—typically augmented with attention mechanisms or merged with other models in ensemble or multimodal scenarios. Research like that of Hameed et al. [22] and Dang et al. [19] also illustrates that the combination of wireless sensing data with deep learning, especially CNNs, yields promising outcomes for emotion tracking. These models tend to outperform conventional machine learning algorithms in both accuracy and robustness to noisy or multimodal inputs.

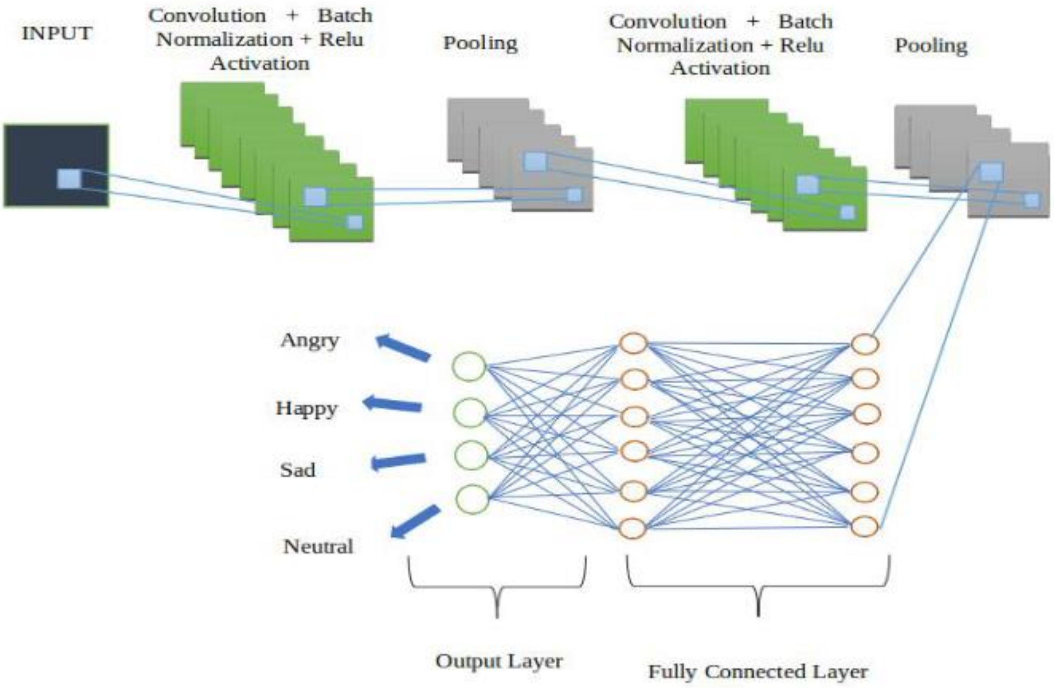

Figure 6: Convolutional Neural Network [23]

2.9 Limitations and Future Work

Wi-Fi Channel State Information (CSI)-based Human Activity Recognition (HAR) and Emotion Recognition (ER) systems have numerous advantages, but there are some challenges to be overcome. The most significant challenge is environmental sensitivity. Minor changes in the environment, for example, rearranging furniture or introducing obstacles, have a significant impact on Wi-Fi signals, which makes it hard to ensure consistent performance across various environments.

Model generalization is another challenge. Getting good recognition across various users and environments is challenging. As Zhao et al. and Gu et al. noted, emotion recognition is particularly challenging since each individual responds uniquely, and such behavioral differences may result in fluctuations of Wi-Fi signals, thus making it more challenging to develop a model that is effective across all people [3][10].

Most existing systems also depend on controlled testing setups, which restrict their usefulness in noisy real-world conditions in which variables such as changing Wi-Fi signals or random human motion can introduce errors.

Moreover, although Wi-Fi CSI-based systems provide superior privacy compared to vision-based systems, there remains a privacy issue. Even when there are no direct images or videos, the Wi-Fi signal can provide insight into an individual's actions or feelings. This is an ethical issue, particularly when users have not given specific consent to being tracked.

Chapter 3. Proposed Methodology

3.1 Introduction

The proposed emotion recognition system offers an inventive method which makes use of Wi-Fi signals for unobtrusive emotion sensing. In a Wi-Fi configuration with transmitter (TX) and receiver (RX), Channel State Information (CSI) is recorded based on faint facial changes reflecting five principal expressions: neutral, happy, sad, fear, and surprise. This information is further processed by feature extraction and separated into training and test sets for machine learning classification. The learning algorithm becomes proficient at categorizing expressions based on signal differences without the need for conventional image-based methods. Classified expressions can be used in diverse real-world applications such as human-robot interaction, biometric authentication, mental state monitoring, and security systems, emphasizing the system's potential in privacy-aware and adaptive emotion perception.

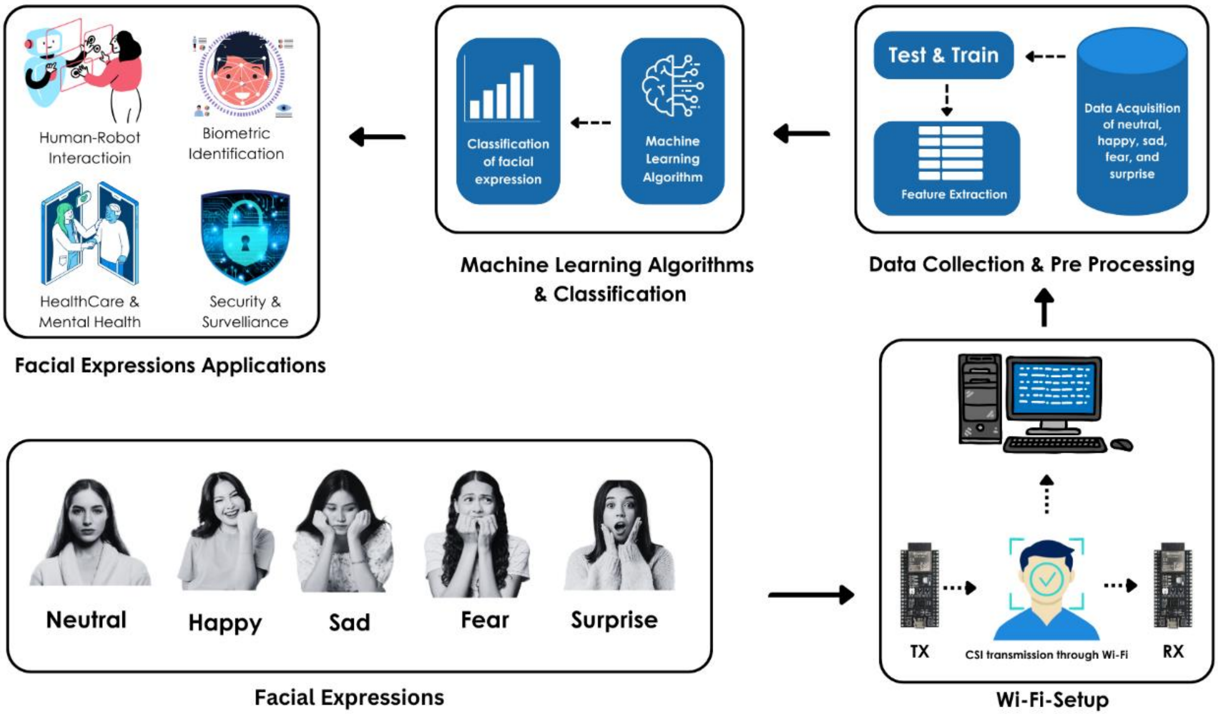

Figure 7: The overall flow diagram of proposed facial expressions system

3.2 CSI Data Collection and Environment Setup

For this purpose, two ESP32 devices were used, One ESP32 was configured in 'Active Station mode' that is used as a transmitter and the second ESP32 operated in 'Active Access Point mode' as a receiver and was connected to a laptop via USB. The CSI receiver collected Wi-Fi signal data transmitted by the sender, which varied based on human movements and gestures in the environment. The tool chain used was based on the ESP32-CSI-Tool repository, which provided support for capturing and decoding CSI packets in real time. CSI data was collected using Visual Studio Code using Python script.

In each experimental session the CSI data that was captured was stored in the .csv file containing approximately 500 CSI packets captured over a period of about 5 seconds. Each row in the CSV represents a packet and includes:

- Timestamp
- RSSI
- Source and destination MAC addresses
- The **CSI_DATA** field, which includes the real and imaginary components of the CSI across multiple subcarriers.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
timestamp	label	type	role	mac	rssi	rate	sig_mode	mcs	bandwidth	smoothing	ot_soundir	aggregatior	stbc	fec_coding	sgi	noise_floor
14:40:16	1	CSI_DATA	AP	10:06:1C:85:A0:10	-72	11	0	0	0	0	0	0	0	0	0	-96
14:40:16	1	CSI_DATA	AP	10:06:1C:85:A0:10	-79	11	0	0	0	0	0	0	0	0	0	-96
14:40:16	1	CSI_DATA	AP	10:06:1C:85:A0:10	-72	11	0	0	0	0	0	0	0	0	0	-96
14:40:16	1	CSI_DATA	AP	10:06:1C:85:A0:10	-72	11	0	0	0	0	0	0	0	0	0	-96
14:40:16	1	CSI_DATA	AP	10:06:1C:85:A0:10	-72	11	0	0	0	0	0	0	0	0	0	-96
14:40:16	1	CSI_DATA	AP	10:06:1C:85:A0:10	-72	11	0	0	0	0	0	0	0	0	0	-96
14:40:16	1	CSI_DATA	AP	10:06:1C:85:A0:10	-72	11	0	0	0	0	0	0	0	0	0	-96
14:40:17	1	CSI_DATA	AP	10:06:1C:85:A0:10	-72	11	0	0	0	0	0	0	0	0	0	-96
14:40:17	1	CSI_DATA	AP	10:06:1C:85:A0:10	-72	11	0	0	0	0	0	0	0	0	0	-96
14:40:18	1	CSI_DATA	AP	10:06:1C:85:A0:10	-72	11	0	0	0	0	0	0	0	0	0	-96
14:40:18	1	CSI_DATA	AP	10:06:1C:85:A0:10	-72	11	0	0	0	0	0	0	0	0	0	-96
14:40:18	1	CSI_DATA	AP	10:06:1C:85:A0:10	-72	11	0	0	0	0	0	0	0	0	0	-96
14:40:18	1	CSI_DATA	AP	10:06:1C:85:A0:10	-72	11	0	0	0	0	0	0	0	0	0	-96
14:40:18	1	CSI_DATA	AP	10:06:1C:85:A0:10	-80	11	1	6	1	1	1	0	0	0	0	-96

Figure 8: Sample of CSV Data

3.3 Data Preprocessing and Spectrogram Generation

3.3.1 Packet Parsing

To analyze the Channel State To obtain preprocessed Channel State Information (CSI) values for emotion recognition, several levels of preprocessing are applied to extract relevant features from the raw signals. Each CSI packet acquired by the ESP32 receiver contains a set of complex values representing how the wireless signal behaves across multiple subcarriers; they represent the changes in the wireless channel due to the subject's facial expressions and body movement during different emotional expressions. Each complex value has two main components: the amplitude (signal strength), and the phase (angle, or time-shifted shift of the signal).

The CSI data is contained in CSV format wherein each row corresponds to a packet in which many of the CSI readings in that packet have been present. The field in the row which contains a relevant set of complex values will be extracted first and then cleaned. This includes all the characters (square brackets, commas etc) that might otherwise be present in the strings. After these are cleaned the values are separated into individual elements. Each individual element will be processed so as to extract the amplitude and phase components. By this separation we can learn exactly which emotions affect both the strength and the phase behavior of the Wi-Fi signals.

After removing the amplitude and phase values from the data, they are converted into numerical representation and organized into structured arrays or matrices using algorithms such as NumPy to facilitate analysis and training of machine learning models. Also, to ensure the consistency of the data used and its suitability for training model, it is further partitioned into fixed size time windows, wherein each window contains a uniform number of CSI samples and acts as one data instance for the training process. This preprocessing pipeline helps to reduce the raw CSI data into a tidy and structured format that is suitable for training and testing deep learning models for emotion recognition.

3.3.2 Complex Amplitude Extraction

The CSI data recorded from ESP32 is a sequence of complicated numbers that tell us how the Wi-Fi signal is behaving between various subcarriers. Complex numbers consist of two components: a real and an imaginary one. Both the real and imaginary parts together denote both the strength and direction of the signal at a given time.

In the CSI data that is extracted, the real part and the imaginary part are held in an interleaved format. This indicates that the real parts are present at even locations within the data (like index 0, 2, 4, and so forth), whereas the imaginary parts are at odd locations (like index 1, 3, 5, and so on). In order to find out how intense the signal is on every subcarrier, we determine the amplitude, or magnitude, of these complex numbers. The amplitude is calculated using the standard mathematical formula:

$$\text{Amplitude}=\sqrt{\{Real^2 + Imaginary^2\}} \qquad \textbf{(4)}$$

This formula provides one positive value that indicates the overall strength of the signal at a particular subcarrier. By calculating the amplitude of every complex number, we can form a clear picture of how the signal is evolving over time.

These movements are usually caused by human activity or emotion-related movement, so the amplitude of data is highly valuable for identifying various emotions.

3.3.3 Amplitude Matrix Construction

In this step, the amplitude values calculated from the complex CSI data were organized into a 2D matrix format. Each packet contains a vector of amplitudes corresponding to 64 subcarriers, which represent the frequency bins in the Wi-Fi signal.

Over time, as multiple packets are received, their corresponding amplitude vectors are stacked row-wise to form a matrix. This matrix provides a structured view of how the signal amplitude varies across subcarriers and time.

- The x-axis of the matrix corresponds to time, with each row representing a separate packet index.
- The y-axis corresponds to the subcarriers, with each column representing a different subcarrier.
- Each cell value in the matrix represents the magnitude (amplitude) of the signal at a particular subcarrier and time instant.

This matrix structure is essential for feature extraction and training machine learning models, as it preserves both temporal and frequency-domain characteristics of the CSI data.

```
Amplitude Matrix (First 5 Rows):
 [[ 69.85699679   1.           0.           0.           0.
    0.          22.8035085   22.36067977  22.8035085   23.34523506
   20.24845673  20.88061302  19.6977156   20.61552813  19.6977156
   21.63330765  21.09502311  18.86796226  20.80865205  20.59126028
   20.          20.61552813  18.60107524  19.41648784  17.4642492
   16.1245155   16.1245155   17.49285568  16.15549442  15.8113883
   17.          18.02775638   0.          17.08800749  16.15549442
   16.76305461  19.31320792  20.88061302  18.68154169  19.23538406
   17.4642492   21.09502311  17.4642492   24.33105012  16.4924225
   20.09975124  19.6468827   19.23538406  20.09975124  22.20360331
   19.23538406  22.20360331  21.          19.23538406  21.09502311
   23.08679276  23.08679276  20.02498439  23.08679276   0.
    0.           0.           0.           0.        ]
 [121.16517652  10.           0.           0.           0.
    0.          19.10497317  19.79898987  20.51828453  21.9317122
   21.9317122   20.          21.40093456  18.86796226  19.72308292
   17.88854382  20.59126028  21.09502311  20.24845673  17.88854382
   17.          18.35755975  17.88854382  17.          16.1245155
   15.26433752  15.8113883   17.20465053  14.76482306  16.64331698
   16.64331698  16.1245155    0.          15.62049935  17.69180601
   17.69180601  19.20937271  16.40121947  16.40121947  20.51828453
   17.80449381  18.43908891  14.2126704   18.43908891  18.43908891
   19.41648784  18.60107524  18.60107524  19.10497317  20.24845673
   18.86796226  19.20937271  19.23538406  20.24845673  19.72308292
   19.41648784  21.09502311  19.23538406  19.72308292   0.
```

Figure 9: Sample amplitude matrix (first 5 packets × 64 subcarriers)

3.3.4 Spectrogram Visualization

To transform the raw CSI amplitude data into a visual representation that is appropriate for deep learning, spectrograms were created using the matplotlib library in Python. Spectrograms give a time-frequency representation of the signal and show how the signal amplitude changes over time. Each spectrogram was produced from an amplitude matrix that recorded the dynamic nature of Wi-Fi signals while performing particular emotional expressions.

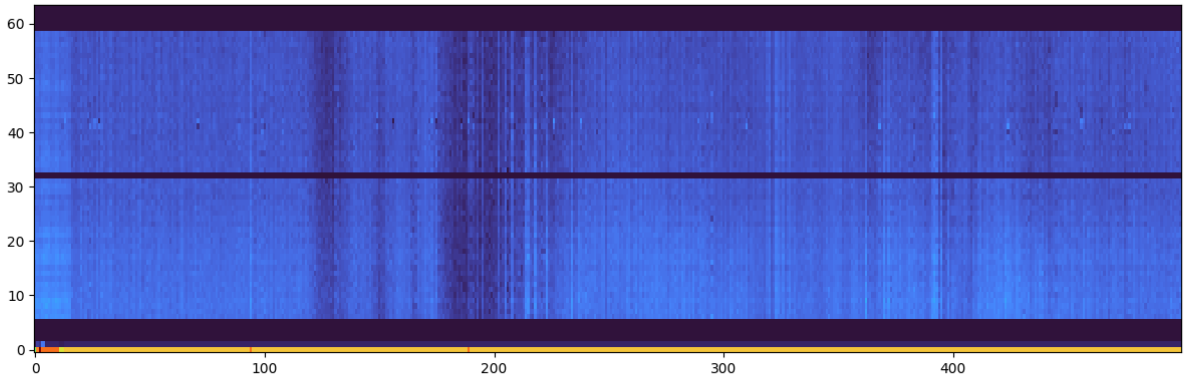

Figure 10: Sample Spectrogram

Figure 10 shows a sample spectrogram image. Each emotional state, e.g., Neutral, Happy, Sad, Fear, or Surprised, creates different patterns in the spectrogram resulting from slight facial muscle movements and physiological reactions. These micro-movements alter the surrounding wireless environment and hence induce small, but measurable changes in the Channel State Information (CSI).

These spectrograms efficiently represent these variations as visual textures, making them very suitable as inputs to image-based deep learning models. In this research, the created spectrogram images were utilized to train a Convolutional Neural Network (CNN). This model was trained to recognize human emotional expressions by learning the characteristic spectrogram features corresponding to each emotional state, illustrating the potential of Wi-Fi CSI as a contactless emotion recognition method.

3.4 Model Description

This section describes the methodology, design choices, and training strategies employed to develop a deep learning model for classifying human emotional expressions into five categories: Fear, Happy, Sad, Surprised, and Neutral. The classification task is based on spectrogram images generated from Wi-Fi CSI data. These spectrograms capture the unique signal patterns associated with each emotional expression and serve as input to a convolutional neural network (CNN) trained using supervised learning techniques.

3.4.1 Dataset Collection and Labeling

A total of 810 spectrogram images were collected from 162 individual participants, with each individual performing five distinct emotional expressions corresponding to the predefined emotion categories. This ensured that there was one spectrogram image per emotion per individual, resulting in a balanced dataset with 162 samples for each emotion class. Each spectrogram image was manually labeled according to the emotion it represented.

The labels were encoded numerically for use in model training, as shown below:

Emotion	Label
Fear	1
Happy	2
Neutral	3
Sad	4
Surprise	5

Table 2: Emotion Categories and Their Corresponding Labels

In order to be organized and allow tracking of every sample, an orderly file naming scheme was used. Every filename started with a prefix that was the participant ID and the respective emotion. For instance:

i. 1F is the spectrogram of the first person displaying the Fear expression.
ii. 2H is the second person's Happy expression.
iii. 159N is the 159th person's Neutral expression.

This naming convention facilitated easily referencing both the subject and its related emotion for each image.

To train the CNN model, a CSV file was created to be used as an input reference to the dataset. The file had two columns:

i. file_name: File name of the spectrogram image file (e.g., 1F.png, 2H.png).
ii. label: Numerical emotion label assigned to each image (e.g., 1 for Fear, 2 for Happy, etc.).

This organized dataset facilitated effective data loading, preprocessing, and label assignment during model training and testing.

3.4.2 Data Splitting Strategies

To provide unbiased and consistent assessment of the deep learning model, there was a systematic data splitting method used. Rather than randomly selecting individual images into training, validation, and test sets, splitting was done across participant identity. This implies that all five spectrogram images corresponding to a particular individual (one for each of the emotions) were placed solely in just one of the three sets: training, validation, or testing.

The data was divided as follows:

i. 80% of the subjects were utilized for the training set
ii. 10% were utilized for the validation set
iii. The other 10% were utilized for the testing set

This was done to prevent data leakage, where multiple images of the same individual are included in both training and test sets. If that were the situation, the model could learn to identify the individual rather than the true emotional patterns, which would result in overly optimistic accuracy and bad generalization.

For instance, if participant 10's images—10F.png, 10H.png, 10sad.png, 10surprised.png, and 10N.png—are included in the training set, none of them can be included in the validation or test sets. This rigid separation guarantees that the model is tested on entirely unseen subjects, which makes the evaluation more realistic and meaningful.

By adopting this identity-based split approach, the learned model generalizes more effectively to novel users, which is particularly crucial for practical emotion recognition systems.

file_name	label
80sad.png	3
35N.png	5
50N.png	5
56N.png	5
13N.png	5
104sad.png	3
57H.png	2
81N.png	5
34S.png	4
59H.png	2
115F.png	1
117N.png	5
8H.png	2

Table 3:Sample CSV file demonstrating two columns

3.5 Initial Experiments

3.5.1 Experiment 1 (ResNet18)

We began our experiments with “ResNet18” architecture due to its ease of use and quick training. Although ResNet18 is reported to be effective on large datasets, it struggled with our small and noisy spectrogram dataset consisting of 810 samples from 162 people (five emotion classes per individual). In the training process, the model exhibited extremely high accuracy — about 98% on training data and 97% on validation data. But the accuracy on the test fell sharply to only 49%. This extreme difference between the training and the test performance hinted that the model was overfitting. It was, in other words, memorizing training instances rather than learning features likely to perform on new, unseen data.

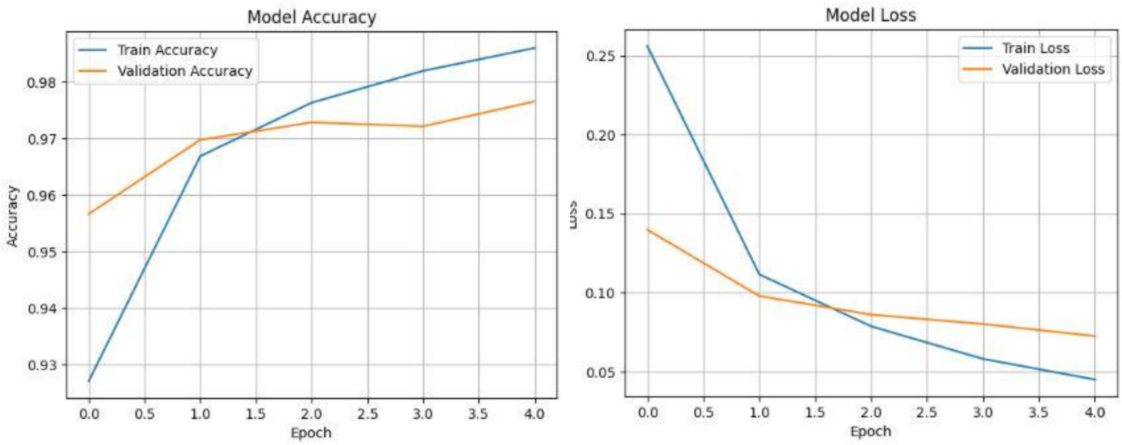

Figure 11: Accuracy and Loss Curves

3.5.2 Experiment 2 (EfficientNetV2B0 & Spectrogram Slicing)

To address the overfitting problem seen in ResNet18, we opted for “EfficientNetV2B0”—a newer CNN architecture geared towards improved performance on small and medium-sized datasets. Its utilization of compound scaling and regularization strategies assisted in increasing the model's learning ability without significantly enhancing overfitting risk. To promote further generalizability, data augmentation was implemented to add variability to the training data and assist the model in learning more generalized features.

Even with these modifications, the results remained unsatisfactory. Examining the data further, we discovered that numerous spectrograms were noisy and had unstructured patterns, which increased the difficulty of distinguishing emotional classes. We overcame this by using a spectrogram slicing technique, where a spectrogram was divided in half along the middle and the right half was positioned below the left half. This reorganization brought far but possibly related features closer, enabling the model to identify meaningful patterns better. Combined with EfficientNetV2B0, this preprocessing process enhanced training and validation accuracy. This result highlighted the significance of not just selecting the appropriate model but also thoroughly prepping the input data for emotion classification tasks on spectrograms.

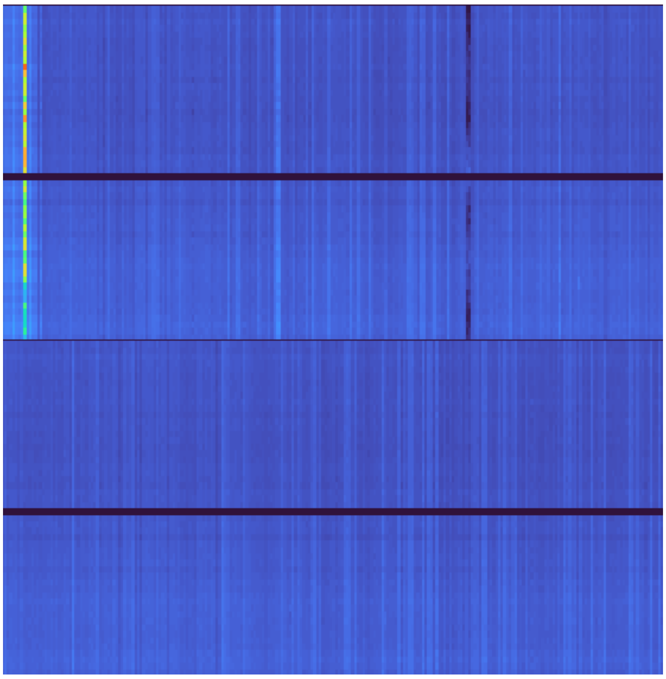

Figure 12: Sliced Spectrogram sample

3.5.3 Experiment 3 (Impact of Data Augmentation)

Data augmentation is a common method to artificially increase the size of the training dataset through random transformations—like rotation, flipping, scaling, cropping, or adding noise—of the initial data samples. This method enhances the diversity of the dataset without needing extra data collection. In this instance, spectrograms were augmented to mimic variations in emotional signal patterns, allowing the model to learn robust and generalized features. In this way, augmentation reduced overfitting—where the model performs very well on training data but cannot generalize to new data.

To see the impact of augmentation on the performance of a model, we trained EfficientNetV2B0 with and without augmentation and compared the training and validation accuracy and loss curves. Without augmentation, the model rapidly reached high training accuracy but had a pronounced dip in validation accuracy, illustrating overfitting. Conversely, using augmentation reduced the gap between training and validation accuracies considerably. This reflected better generalization, since the model was now being exposed to a greater variety of input variations during training. The validation loss curve also became smoother and more stable, reflecting that the model was learning more consistent and reliable features when trained on augmented spectrograms.

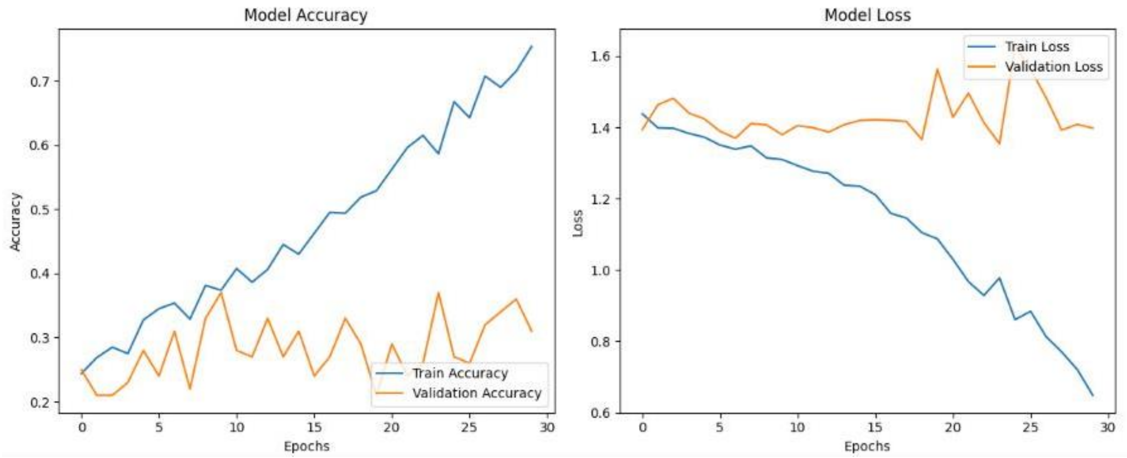

Figure 13: Slicing without augmentation result

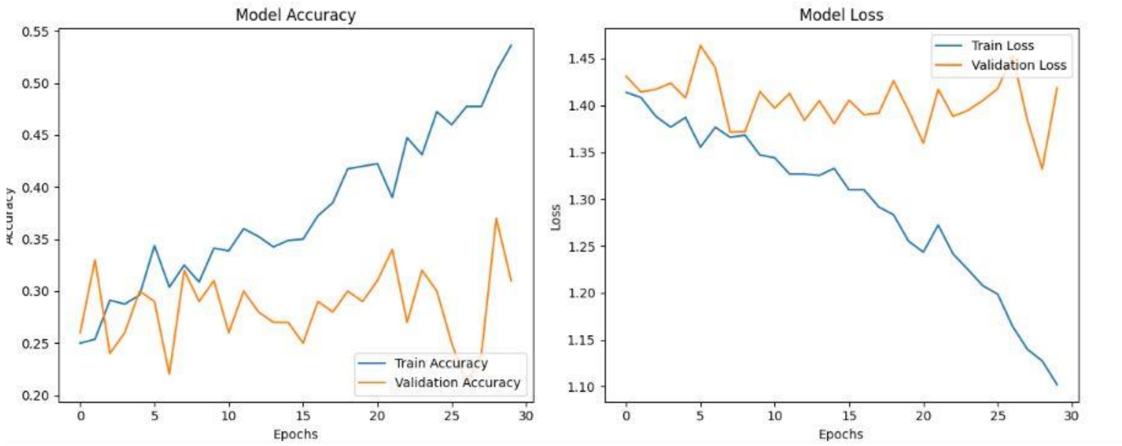

Figure 14: Slicing with Augmentation result

3.5.4 Experiment 4 (Absolute difference)

Following on our attempts to better model performance with EfficientNetV2B0, we discovered that slicing spectrograms alone, while beneficial, was not enough to improve significantly the model's capacity to generalize. Although the slicing method bettered spatial relations by bringing varying parts of the spectrogram together, it still did not remedy the problem of noise in the original spectrograms. Most of the raw spectrograms had irrelevant or overlapping frequency components, which made it difficult for the model to concentrate on the real emotional features.

In order to do this, we added a new preprocessing step: computing the absolute difference between each emotional spectrogram and its corresponding Neutral spectrogram for the same person. For instance, we did a subtraction of the Neutral spectrogram of person 5 (5N) from their Fear (5F), Happy (5H), Sad (5S), and Surprised (5Su) spectrograms, yielding better images named like 5N-F, 5N-H, 5N-S, and 5N-sad. This maneuver highlighted motion and energy pattern differences, essentially extracting the characteristics peculiar to each emotion by filtering out shared background noise recorded in the Neutral state.

We then applied our slicing method to these difference images. This fusion not only retained meaningful features but also increased contrast between emotional expressions, facilitating the EfficientNetV2B0 model to learn class-specific patterns more easily. Consequently, we noticed a significant improvement in validation and test accuracy, validating that handling data quality through noise reduction was as important as model choice. This process reaffirmed the finding that effective preprocessing according to the type of data can contribute more than deeper models or data augmentation by themselves.

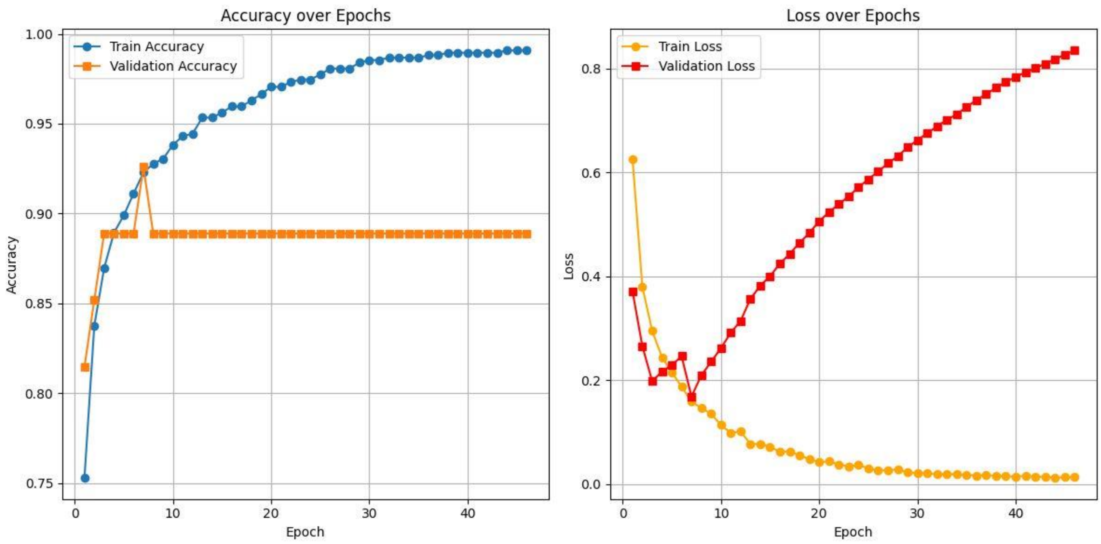

Figure 15: Absolute Difference with slicing & without augmentation

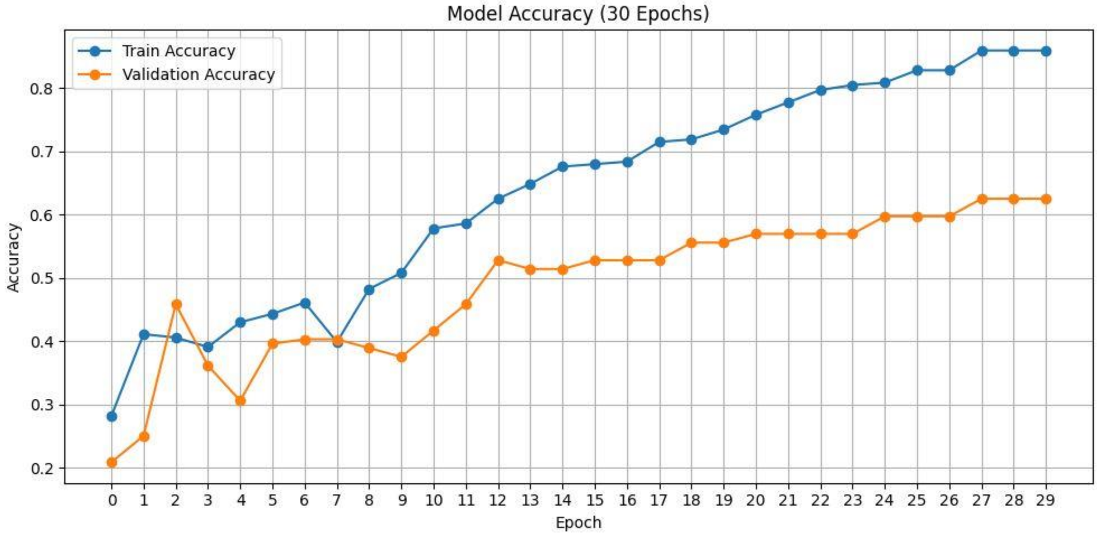

Figure 16: Absolute Difference with slicing & with augmentation

A comparative analysis was conducted to assess the effect of data augmentation on model generalization. The two models, EfficientNetV2B0 with and without augmentation, were trained with the same dataset containing training, validation, and test splits. The first model, however, used data augmentation methods like random shifts and scaling to artificially introduce diversity in training data, while the second model did not use any.

When both models were trained, the performance of the two was identical at the beginning but diverged with the progression of training. The augmented model had much better validation accuracy (about 75–80%) than the non-augmented model, which stagnated at about 45%. Although the non-augmented model had more than 99% training accuracy, its validation loss was still high and fluctuating, which refers to bad generalization to out-of-sample data.

This behavior exhibits classic overfitting—where the model has memorized training data but not learned generalizable meaningful patterns. By contrast, the augmented model had smoother and more stable loss plots and a reduced difference between training and validation accuracy. These outcomes verify that data augmentation significantly helps combat overfitting and improve the robustness of emotion classification models learned from spectrogram data.

Chapter 4. Hardware Implementation, Testing and Results

For collecting Wi-Fi Channel State Information (CSI) data, two ESP32 development boards were utilized, one functioning as a transmitter and the other as a receiver. These devices were configured using the ESP32 CSI Tool firmware, which enables extraction of fine-grained CSI data from Wi-Fi packets.

Figure 17: ESP 32

4.1 Hardware Configuration

- **Transmitter (Station Mode):**

One ESP32 board was configured in Station Mode. In this mode, it actively transmitted continuous Wi-Fi packets towards the receiver. This ESP32 acted as the source of Wi-Fi signals, emulating a real-time transmission environment.

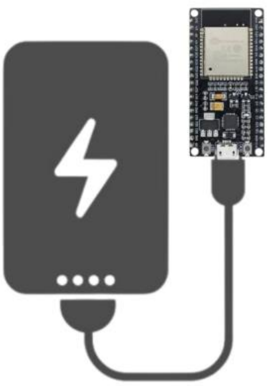

Figure 18: Transmitter ESP32 powered by power bank

- **Receiver (Access Point Mode):**

  The second ESP32 was configured in Access Point Mode. It served as the receiver, capturing the incoming Wi-Fi packets transmitted by the Station-mode ESP32. From these packets, the board extracted the CSI values, which were then transferred via USB connection to a laptop for real-time storage and processing.

Figure 19:Receiver ESP32 connected to laptop (with Reflector)

The ESP32 modules were strategically placed to ensure that reflected signals from the participant's facial region were captured with maximum amplitude variation. This setup aimed to enhance the expressiveness of the received CSI data, improving the model's ability to learn subtle emotional cues. The goal was to enable the CNN model to train with high accuracy and low validation loss by maximizing the quality of input data.

- **Power Setup**

To support uninterrupted operation during data collection:

i. The Receiver ESP32 (Access Point Mode) was powered through the USB connection to the laptop, which also handled data storage and further processing.

ii. The Transmitter ESP32 (Station Mode) was powered using a portable power bank, providing a stable and consistent power supply without reliance on wired connections. This configuration also ensured mobility and isolation from the laptop to minimize electromagnetic interference.

4.2 Environment for CSI Data Collection

The setup for CSI data acquisition was carefully crafted to maximize signal clarity and reduce interference. The transmitter and receiver ESP32 boards were set at specific angles to enable Wi-Fi signals to bounce directly off the participant's facial area and hit the receiver with the greatest possible amplitude. This deliberate positioning was essential in capturing the fine signal variations resulting from micro-movements involved in various facial expressions, which are evident in the spectrograms.

To preserve data integrity, it was guaranteed that no further movement took place in the environment around the data recording. The participant was the only one who could move, essentially reducing noise and avoiding unwanted fluctuations in the CSI data. Moreover, all data was obtained within an enclosed room to avoid any external interference from wireless sources and signal loss. This controlled environment served a crucial purpose in improving the quality of CSI measurements and towards the better accuracy and reliability of the emotion recognition model.

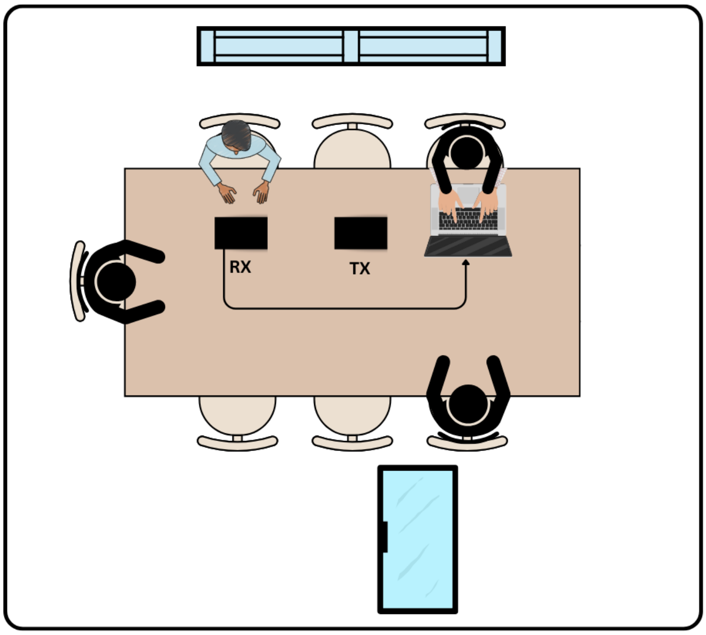

Figure 20: Experimental Environment

The setup depicted indicates the controlled setup employed for taking Wi-Fi CSI data for recognizing facial expressions. An individual sits in front of two ESP32 boards on a table — one set up as a Transmitter (TX) and the other as a Receiver (RX). The TX continually transmits Wi-Fi signals, which bounce back from the individual's face and are received by the RX. The receiver is connected to a laptop using USB, where the CSI data is recorded in real time. This configuration has the effect of ensuring that even minor changes in facial expressions bring about perceivable variations in received signals.

The setup is then put inside a closed room in order to preclude external interference. Members from the team stand by to ensure the process runs smoothly and there is no other movement during the process of gathering data. The machines are set at specific angles in such a way that the face-reflected signals hit the receiver with the greatest possible strength, thereby contributing to clearer spectrograms. This setup is made to enhance data quality and aid in improved model training and accuracy.

The CSI data was tracked and fetched with the aid of Visual Studio Code alongside the ESP-IDF CSI tool. The ESP32 receiver module, which was in Access Point (AP) mode, recorded raw Channel State Information (CSI) packets upon transmission. Such packets were diverted to the accompanying laptop, where they were stored in .csv (Comma-Separated Values) format for post-processing.

A single `.csv` file comprised comprehensive information about the recorded packets, such as:

i. Timestamps that tell exactly when every packet was received,
ii. MAC addresses of both the receiving (source) and sending (destination) ESP32 devices, and
iii. RSSI (Received Signal Strength Indicator) values that are measures of the power of the received signal.

This formatted data provided the basis for creating spectrograms and for training machine learning models.

4.3 Spectrogram Samples

The CSI data obtained from the receiver ESP32 was saved straight into a .csv file while collecting the data. For every test case, there was a 5-second time window within which the transmitter and receiver ESPs formed a communication link. In this window, the user made a particular facial expression (e.g., Neutral, Happy, Sad, Fear, or Surprised), and the respective raw CSI packets were recorded. For every file that logged necessary information, it noted the timestamps, destination and source MAC addresses, as well as the signal parameters of the specific expression.

For creation of the spectrograms out of this information, a dedicated Python script was crafted. It mined the CSI signals' real and imaginary values out of the.csv file, calculated their amplitudes, and proceeded to plot the results as a matrix of sub-carrier-amplitudes across time. The spectrograms obtained contained meaningful visual information about CSI data and were used as inputs to machine learning models. The precise steps of this conversion are further discussed in the Modeling and Simulation section of this thesis.

4.3.1 Spectrograms

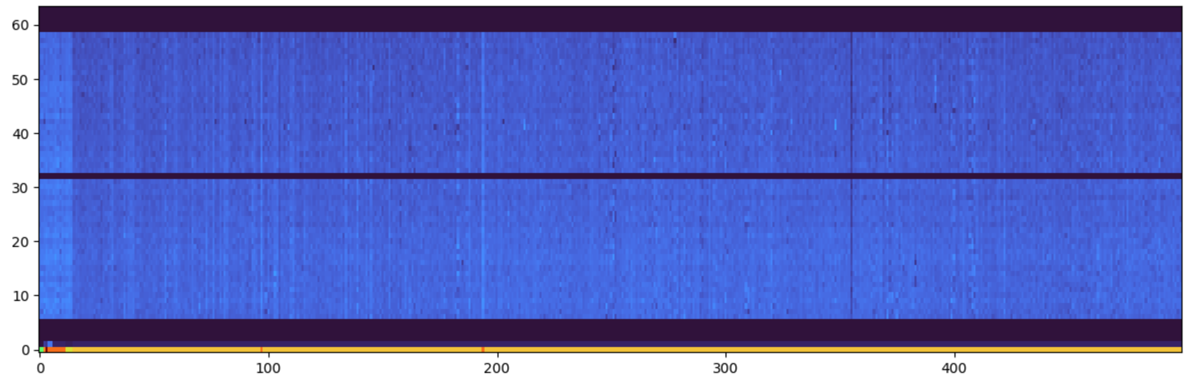

Figure 21: Spectrogram of Neutral expression

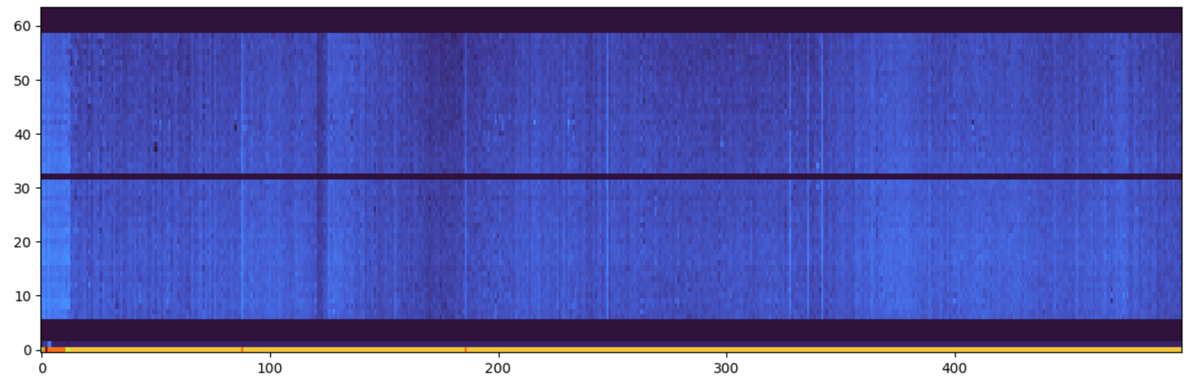

Figure 22: Spectrogram of Fear expression

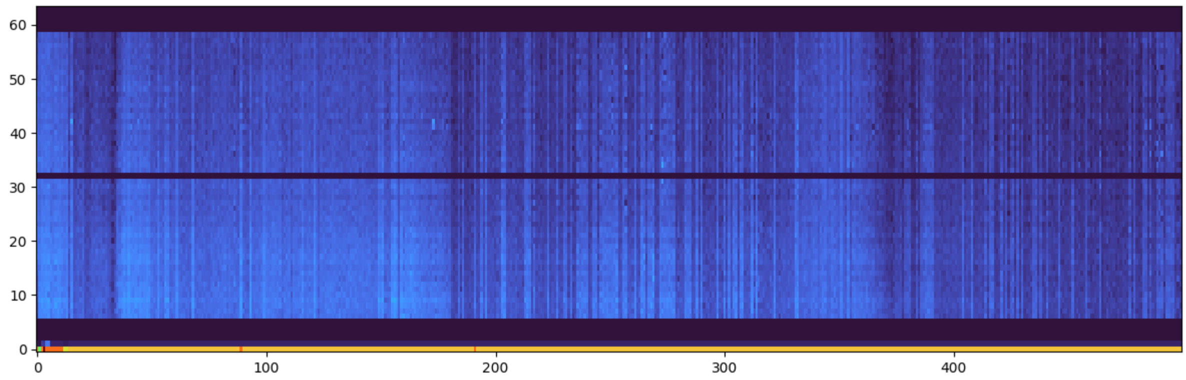

Figure 23: Spectrogram of Sad expression

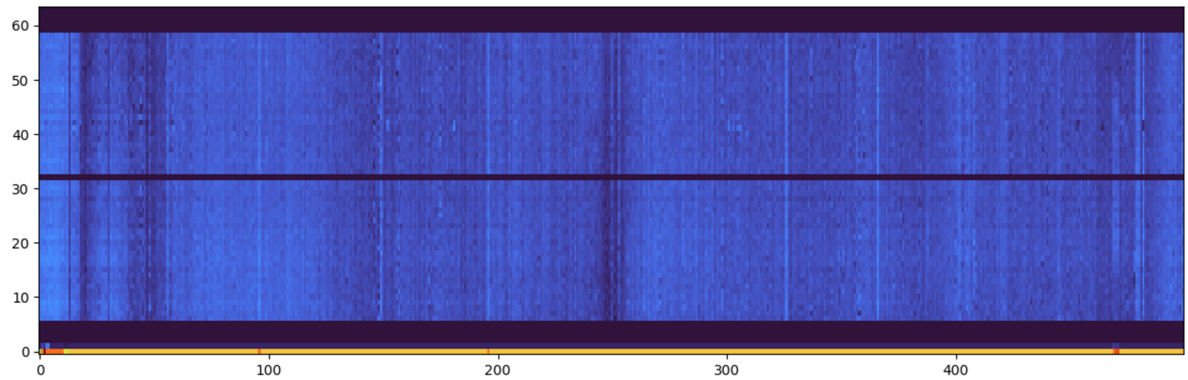

Figure 24: Spectrogram of Happy expression

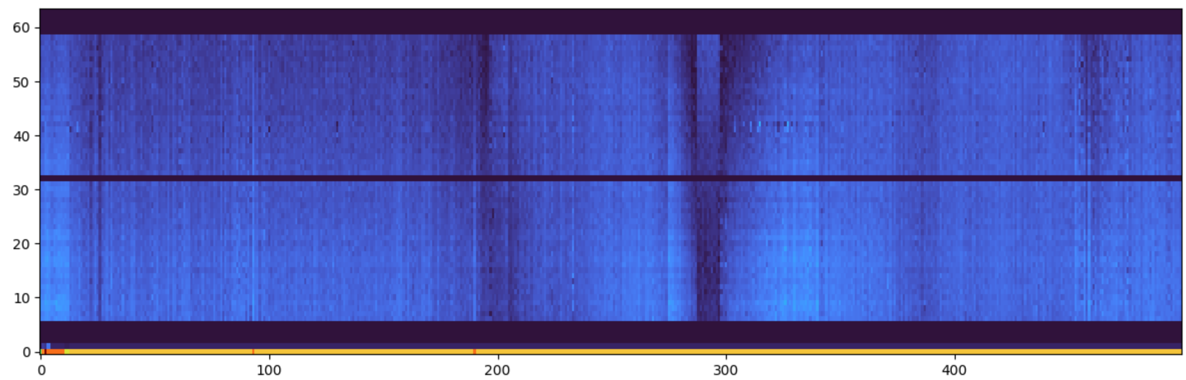

Figure 25: Spectrogram of Surprised expression

It can clearly be observed that the neutral signal depicts very small variations in amplitude across subcarriers, signaling minimal movement in the facial muscles and an intact expression over the window of recording.

Chapter 5. Initial Results

Through improved data processing and analysis techniques, the model saw notable enhancements. The addition of a reflector to improve output clarity by strengthening signal reflections was one of the main improvements. To ensure that no important features were lost during preprocessing, full RGB images were used to maintain enrollment signatures. The dataset was divided into training, validation, and test sets at random to reduce bias and ensure a fair evaluation. Additionally, misclassifications were thoroughly examined to find underlying trends and enhance model functionality.

With training accuracy reaching 90% by the tenth epoch and validation and test accuracy following closely behind 88% and 85%, respectively, accuracy metrics showed consistent improvement over epochs. Although small discrepancies between training and test accuracy point to possible overfitting, this progression demonstrates the model's capacity for effective learning. Earlier iterations of the model used simple spectrograms to extract features, which were helpful but lacked the depth necessary to identify subtle patterns in the data. To increase the dataset, image slicing and augmentations were used, although these methods occasionally resulted in distortions. Separate CSV files were used for data management at first, which increased the possibility of overlap and inconsistency. Additionally, targeted improvements were hampered by the preliminary phases' lack of systematic mistake tracking.

In conclusion, the model's robustness was improved by the refinements, which included randomized splitting, error analysis, and optimized data inputs. Future iterations will be guided toward greater precision by the limitations of early approaches, which highlighted the significance of complex feature extraction and reliable error tracking.

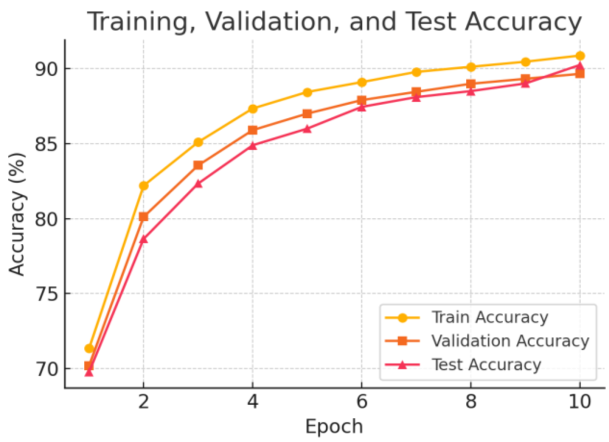

Figure 26: Accuracy Curve

Chapter 6. Conclusion and Future work

6.1 Conclusion

This work has investigated the application of Wi-Fi Channel State Information (CSI) for emotion and activity recognition, showcasing the potential of RF-based sensing as a non-intrusive, privacy-respecting, and cost-efficient substitute for conventional camera or wearable-based systems. By data collection, preprocessing, spectrogram generation, and training deep learning models, we showed that changes in Wi-Fi signals can effectively detect and classify human emotions and activities with promising accuracy. The method is especially well-suited for use in applications where user privacy, comfort, or environmental restrictions prevent the use of visual or physical sensors.

The system built in this work adds to the expanding literature on wireless human sensing, demonstrating that common communication signals can be reused for intelligent, human-aware interaction. It also provides a foundation for contactless monitoring and control applications in smart homes, healthcare, transportation, and public spaces.

6.2 Real-Time CSI Labeling Application and Future Work

As a key outcome of this project, we developed a mobile application—CSI Labeller—to perform real-time emotion recognition using Wi-Fi Channel State Information (CSI). This Android app functions both as a data labeling tool and a complete inference system, combining data collection, processing, and classification in a single user-friendly interface.

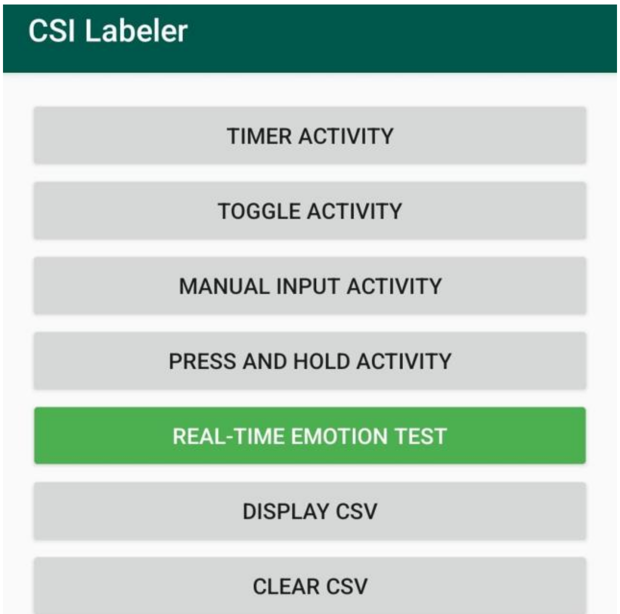

Figure 27: User Interface

6.2.1 Core Features and Implementation

Real-time CSI data collection from ESP32 Wi-Fi modules is made possible by the CSI Labeller application, and the data is directly stored on the device in CSV format. A pretrained deep learning model (ResNet18), which is integrated into the application using PyTorch Mobile, processes the spectrogram images created from this CSI data. The following essential features are part of the app's architecture:

1. Offline Inference Capability: The application comes with a TorchScript version of the trained model (.pt format), which enables on-device emotion classification without an internet connection.
2. Live Emotion Recognition: Continuous emotion detection from streaming CSI data is made possible by a special "Real-Time Emotion Test" module.
3. Prediction Feedback: The program offers interpretable insights into the detected emotion, related prediction confidence, and corresponding spectrogram in real-time.

6.2.2 Usability and Performance Evaluation

The app is built with a user-friendly and interactive user interface, with touch-based controls for live inference as well as manual testing. Emotions are easily classified, and results are provided within seconds, so there's minimal lag and smooth user experience. On-device inference greatly improves responsiveness, and the app is therefore ideal for real-time use. Moreover, the modular design of the app enables it to operate as both a data capture and assessment platform. This two-in-one feature supports iterative refinement in model training and performance evaluation, making it easier to develop the CSI-based emotion recognition system. The major benefits it provides are:

1. Privacy-Preserving Design: Being Wi-Fi signal-based only, the CSI Labeler does not capture visual data, thus maintaining user privacy.
2. Real-Time Processing: Utilizing lightweight neural network models and mobile hardware, the application allows for fast processing appropriate for time-critical situations.
3. End-to-End Integration: From CSI capture to spectrogram conversion and model inference, all processes are consolidated in the application for easy experimentation and deployment.
4. Portable and Cost-Effective: Based on low-cost ESP32 devices and common Android smartphones, the solution is scalable and available across a variety of use cases.
5. Explainable Output: Visual outputs in the form of predicted emotions and spectrograms improve transparency and usability for both researchers and end users.

6.2.3 Future Applications in IoT

As part of the continued evolution of wireless sensing systems, our future research will involve extending the functionalities of emotion and activity recognition through the incorporation of other human-centered applications. The aim is to investigate the possibility of Wi-Fi-based sensing for real-time, non-intrusive interaction and monitoring across various environments.

Facial Recognition Using Wi-Fi Signals

We propose to deploy facial recognition based on Channel State Information (CSI) in order to identify people without having cameras. Privacy-preserving and applicable for smart homes, secure access systems, and personalized services where identification of users is paramount.

Hand Gesture Recognition

Future advancement is to identify hand gestures via variations in wireless signals by hand movements. This would allow for natural, touch-free interaction with systems and devices, which are most useful in settings where physical contact is inconvenient or restricted.

Sign Language Recognition

Leveraging gesture recognition, we envision creating a real-time sign language recognition system through RF signal changes. This is significant for accessibility, allowing continuous communication assistance to the deaf and hard-of-hearing population.

Emotion Recognition in Real-Time Using Wi-Fi Signals

We will increase the system capability to conduct real-time emotion identification through the investigation of fine-grained signal modification associated with bodily movements and face expressions. Such can be enormously beneficial in intelligent healthcare, psychological well-being surveillance, and accommodating user interfaces.

Driver Fatigue and Distraction Detection

One of the more promising areas is the application of Wi-Fi signals to track driving habits for indications of drowsiness or distraction. With the ability to detect minute changes in posture, head movement, and responsiveness, the system might be used as an early alert system to improve road safety.

Touchless Control Interfaces for Public or Medical Use

We also plan to create gesture-based, touchless interfaces driven by wireless sensing for public kiosks, hospitals, and clean room applications. These interfaces will minimize contamination risk and enhance user interaction in sensitive settings.

6.3 Suggestions for Improving the Proposed System

In order to enhance the performance and scalability of the system as of now, we suggest a number of improvements. Firstly, scaling up the dataset size and variety would enable the model to generalize more across different subjects and contexts. Secondly, trying out deeper neural networks like attention-based architectures (e.g., Transformers) or hybrid CNN-RNN architectures would be able to extract spatial and temporal features more efficiently.

At the hardware level, employing multiple ESP32 receivers or more sensitive receivers would enhance signal quality and reliability. Real-time deployment on embedded systems can also be optimized with model quantization or lightweight models such as MobileNet. Last but not least, creating an easy-to-use user interface, particularly a mobile app—for real-time visualization and feedback would make a big difference in usability and user interaction.

6.4 Final Thoughts and Recommendations

This thesis shows the feasibility of RF-based emotion and activity recognition using Wi-Fi signals and is an initial step toward more pervasive human sensing systems. As wireless infrastructure grows and becomes increasingly pervasive, its potential for intelligent, contactless interaction will grow exponentially.

We propose that future research prioritize implementation in real-time systems, ethical design around privacy, and interdisciplinary collaborations to introduce RF sensing solutions into general applications. With ongoing innovation and development, Wi-Fi-based sensing has the ability to transform the way humans interact with technology—securely, transparently, and cognitively.

Chapter 7. Bibliography

[1]. Pise, A. A. et al., 2022. Methods for facial expression recognition with applications in challenging situations. Computational Intelligence and Neuroscience, Article ID 9261438.

[2]. Tao, W. et al., 2023. EEG-Based Emotion Recognition via Channel-Wise Attention and Self Attention. IEEE Transactions on Affective Computing, 14, pp. 382–393.

[3]. Gu, Y. et al., 2020. WiFE: WiFi and Vision based Intelligent Facial-Gesture Emotion Recognition. arXiv preprint arXiv:2004.09889.

[4]. Hou, Y. et al., 2022. WiFi and Vision enabled Multimodal Emotion Recognition. In: Proceedings of the IEEE International Conference on Communications (ICC), pp. 3034–3039.

[5]. Li, H. et al., 2018. Wi-Motion: A Robust Human Activity Recognition Using WiFi Signals. IEEE Access, 7, pp. 153287–153299.

[6]. Wu, Y., Zhang, L., Gu, Z., Lu, H. & Wan, S., 2023. Edge-AI-driven framework with efficient mobile network design for facial expression recognition. ACM Transactions on Embedded Computing Systems, 22, pp. 1–17.

[7]. Komagal, E. & Yogameena, B., 2023. PTZ-camera-based facial expression analysis using Faster R-CNN for student engagement recognition. In: Computer Vision and Machine Intelligence Paradigms for SDGs: Select Proceedings of ICRTAC-CVMIP 2021. Springer, pp. 1–14.

[8]. Dzedzickis, A., Kaklauskas, A. & Bucinskas, V., 2020. Human emotion recognition: Review of sensors and methods. Sensors, 20(3), p. 592.

[9]. Strohmayer, J. & Kampel, M. Directional Antenna Systems for Long-Range Through-Wall Human Activity Recognition. Computer Vision Lab, TU Wien, Austria.

[10]. Zhao, M., Adib, F. & Katabi, D. Emotion Recognition using Wireless Signals. MIT CSAIL. [Unpublished technical report].

[11]. Tong, T. V. A., Bui-Thanh, B. & Nguyen, P. T. H. Human Activity Recognition using Wireless Signals and Low-Cost Embedded Devices. University of Information Technology, Ho Chi Minh City, Vietnam.

[12]. Hernandez, S. M. & Bulut, E., 2020. Lightweight and Standalone IoT Based WiFi Sensing for Active Repositioning and Mobility. In: Proceedings of the 21st International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), Cork, Ireland.

[13]. Pham, M. T., Hoang, L. T., Hong, H. D. T., Minh, T. D. & Nguyen, T. H. P., 2024. A WiFi CSI signal enhancement framework for activity recognition using machine learning automatic segmentation. SN Computer Science, 5(5), p. 524.

[14]. Shahverdi, H. et al., 2023. Convolutional Neural Network Based Human Activity Recognition using CSI. International Journal of Information and Communication Technology Research.

[15]. Moshiri, P. F. et al., 2021. CSI-Based Human Activity Recognition using Convolutional Neural Networks. In: Proceedings of the 2021 International Conference on Computer Engineering and Knowledge (ICCKE), pp. 7–12.

[16]. Moshiri, P. F. et al., 2021. A CSI-Based Human Activity Recognition Using Deep Learning. Sensors (Basel, Switzerland), 21.

[17]. El Zein, H. et al., 2024. CSI-Based Human Activity Recognition via Lightweight CNN Model and Data Augmentation. IEEE Sensors Journal, 24, pp. 25060–25069.

[18]. Zhuravchak, A. et al., 2021. Human Activity Recognition based on Wi-Fi CSI Data – A Deep Neural Network Approach. In: Proceedings of EUSPN/ICTH.

[19]. Dang, X., Chen, Z., Hao, Z., Ga, M., Han, X., Zhang, X., Yang, J. Wireless Sensing Technology Combined with Facial Expression to Realize Multimodal Emotion Recognition. Sensors, 2023, 23, 338.

[20]. Li, F., Al-qaness, M.A.A., Zhang, Y., Zhao, B., Luan, X. A Robust and Device-Free System for the Recognition and Classification of Elderly Activities. Sensors, 2016, 16, 2043.

[21]. Zheng Yang, Zimu Zhou, and Yunhao Liu. 2013. From RSSI to CSI: Indoor localization via channel response. ACM Comput. Surv., 46(2), Article 25 (November 2013), 32 pages.

[22]. Hameed H., Elsayed M., Kaur J., Usman M., Tang C., Ghadban N., Kernec J. L., Hussain A., Imran M., Abbasi Q. H. RF sensing enabled tracking of human facial expressions using machine learning algorithms. Scientific Reports, 2024 Nov 13; 14(1):27800. doi: 10.1038/s41598-024-75909-w.

[23]. Nandan, V. G., Shivakumar, S., Sangeetha, J., Nayak, M. P., & Nishanth, S. K., 2021. A Comparative Study of Deep Learning and Machine Learning Approaches in Speech Emotion and Gender Recognition System. Natural Volatiles & Essential Oils, 8(5), 12261–12273.

Document Viewer

Turnitin Originality Report

Processed on: 12-May-2025 4:01 AM PDT
ID: 2673770075
Word Count: 10939
Submitted: 1

Thesis By Saber Jan

Similarity Index	Similarity by Source	
7%	Internet Sources:	5%
	Publications:	4%
	Student Papers:	4%

exclude quoted | exclude bibliography | exclude small matches | mode: quickview (classic) report | print | refresh | download

1% match (Internet from 20-Oct-2023)
https://www.coursehero.com/file/69205451/HWTR-Thesisdocx/

1% match (Internet from 25-Mar-2025)
https://www.coursehero.com/tutors-problems/Electrical-Engineering/52304888--Question-3-15-pts-Consider-the-case-of-the-interior-of-a/

1% match (student papers from 26-Nov-2013)
Submitted to Higher Education Commission Pakistan on 2013-11-26

1% match (Ma, Yongsen. "Improving WiFi Sensing and Networking with Channel State Information", The College of William and Mary, 2020)
Ma, Yongsen. "Improving WiFi Sensing and Networking with Channel State Information", The College of William and Mary, 2020

<1% match (Internet from 26-Dec-2024)
https://www.coursehero.com/file/232339393/Harshithadocx/

<1% match (Internet from 04-Jul-2024)
https://www.coursehero.com/file/159821739/IOT-based-ditribution-transformer-protection-system/

<1% match (student papers from 25-May-2010)
Submitted to Higher Education Commission Pakistan on 2010-05-25

<1% match (student papers from 27-Nov-2013)
Submitted to Higher Education Commission Pakistan on 2013-11-27

<1% match (student papers from 04-Jan-2011)
Submitted to Higher Education Commission Pakistan on 2011-01-04

<1% match (publications)
Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dhirendra Kumar Shukla. "Intelligent Computing and Communication Techniques - Volume 2", CRC Press, 2025

<1% match (Internet from 13-Jan-2025)
https://www.jetir.org/papers/JETIR2403808.pdf

<1% match (Internet from 12-Feb-2025)
https://www.jetir.org/papers/JETIRGA06033.pdf

<1% match (Jesuola, Goodness Dara. "Impact of the Nigerian Government's Fuel Subsidy Removal: Data Analysis and Suggestions", Texas A&M University - Kingsville, 2024)
Jesuola, Goodness Dara. "Impact of the Nigerian Government's Fuel Subsidy Removal: Data Analysis and Suggestions", Texas A&M University - Kingsville, 2024

<1% match (Susanna Mosleh, Jason B. Coder, Christopher G. Scully, Keith Forsyth, Mohamad Omar Al Kalaa. "Monitoring Respiratory Motion with Wi-Fi CSI:Characterizing performance and the BreatheSmart Algorithm", IEEE Access, 2022)
Susanna Mosleh, Jason B. Coder, Christopher G. Scully, Keith Forsyth, Mohamad Omar Al Kalaa. "Monitoring Respiratory Motion with Wi-Fi CSI:Characterizing performance and the BreatheSmart Algorithm", IEEE Access, 2022

<1% match (Internet from 24-Jul-2024)
https://www.mdpi.com/2813-2203/3/3/17

<1% match (Internet from 09-Apr-2018)
http://www.mdpi.com

<1% match (Internet from 20-Jan-2025)