ISSN: 2834-4928

Research Article

Journal of Electrical Electronics Engineering

A Voice-Activated, Deep-Learning-Based Emergency Alarm for Arabic Dialects

Mohammed Z. Al Zahrani, Helal J. Al Malki, Saud S. Al Lahyani, Ali A. Al Zahrani, Khalid O. Al Malki, Ahmad Z. Hasanain[1]*, Mirza I. Yunus[1], Ashraf M. Khalifa[1], Mohammed M. Almatrafi[1]

College of Engineering at al-Lith, Umm Al-Qura University, SaudiArabia

***Corresponding Author**
Ahmad Z. Hasanain, College of Engineering at al-Lith, Umm Al-Qura University, SaudiArabia

Submitted: 2023, Apr 10; **Accepted:** 2023, Apr 24; **Published:** 2023, May 18

Citation: Al Zahrani, Z. M., Al Malki, H. J., Al Lahyani, S. S., Al Zahrani, A. A., Al Malki, K. O,. et. al. (2023). A Voice-Activated, Deep-Learning-Based Emergency Alarm for Arabic Dialects. *J Electrical Electron Eng, 2*(2), 139-167.

Abstract

In this report, we will talk in almost every detail about our project, in order to cover most aspects of the project. We tried as much as possible to make the report suitable for two segments, the specialists and the public so that the reader can make the most of the report.

We will talk about the hardware and software used in the project and how they work, in order to explain to the ordinary reader how some technical matters are done and to explain to the specialist some things that he or she might consider while developing. In our project, samples are the cornerstone of the project, and for this, we have focused on them in the report, from the methods of identifying samples to access to taking samples and analyzing them project to other dimensions. When we wrote the project, we took into account that development on the project is an important part of this academic process, so we paved the way for those who want to develop by collecting our own database in Arabic for scientific purposes so that development on it becomes easier and more productive.

Keywords: Fire Detection, Speech Recognition.

Article Highlights

- In emergency situations, what people need most when they are faced with an emergency situation is a device that communicates their needs to their loved ones.
- We have collected 967 samples. The positive ones are 382, and the number of negative samples is 585. They were collected by the team members from 41 participants of different ages, genders, nationalities, and background voices.
- We have outlined a sound emergency alarm framework with deep learning of the Arabic language and the dialects of Saudi Arabia

List of Abbreviations

ASR – Automatic Speech Recognition
API – Application Programming Interface
BLE – Bluetooth Low Energy
LED – Light Emitting Diode
IDE – Integrated Development Environment
IoT – Internet of Things
3D – Three Dimensional
MIT – Massachusetts Institute of Technology
OSX – On-Screen Keyboard
C++ – a general-purpose programming and coding language
AVR- C – a family of microcontrollers programmed by the C language
RGB – Red Green Blue
hPa – hectoPascal Pressure Unit
SIG – Special Interest Group
UUID – Universally Unique Identifier
MFCC – Mel Frequency Cepstral Coefficients
PDM – Pulse Density Modulation
WAV – Waveform Audio File Format

1. Introduction

Smart devices in our time are among the most important products of modern life. Smart devices have made life easier, and planning things has become easy and fast to make our lives smoother. In light of the technological development that occurs on a daily basis, devices have been developed rapidly, and many of these smart devices have been invented to become an important part of the daily life of all people. Since technology works greatly on changing the way we live, we always need helpers in our daily life, due to our busy schedule. Accordingly, smart devices can help us in many different ways through very simple commands by programming smart devices. Among the advantages of the smart devices that is provided to the users is the advantage of accessing a lot of information by a simple touch. For example, the users can access televisions, computers, and other smart devices in their homes with a single touch on the mobile screen from anywhere in the world or by voice. Smart devices are easy to use; anyone can use them, and even beginners can use smart devices with little knowledge.

Smart devices play a major role in making life more comfortable, so the user can use the voice command while he or she is sitting in his or her place and does not have to go to certain places. This applies to turning on and off the lights and other electronic tools using the voice command. Additionally, due to the lack of time for most people; they cannot visit the doctor on a regular basis. However, with the help of smart health devices, you can get an alert about your health, your health condition, and your next doctor appointment. These devices are capable of tracking a person's movements, such as walking steps, heart rate, and calories. Furthermore, the technology greatly helps in entertaining its users. There are various entertainment methods, such playing the users' favorite music, videos, or the content they prefer on the smart television screens. There are many reasons why we use smart devices, and transform our life and home into life and smart home, the first and last reliance of them on modern technology.

Among the reasons that make smart devices of great importance is efficiency. It will enable you, through voice, to control many tools and systems around you. With the help of the smart device, you'll be able to turn on the heating, cooling, and lights off with your voice from anywhere in your home. Smart devices are a major way to gain peace of mind; you can use your smart device to check the doors and windows and see if there is a water leak or not, and you can also check that the gate of your house is closed through an application on your device. You don't need to be at home to check all this out; you can do that while sitting in your place. You can own electronic things however you wish to have them. You can adjust the brightness of the internal and external lighting according to your choices, and you can customize each electronic element according to your will and can set specific times for the things you want to do and perform them. Regarding water and electricity bills, we always leave the lights on due to laziness standing on and off, but a smart home allows you to turn off the lights and all other electronic devices in the house. It will save you a large amount of money in light of the fact that the world is getting more expensive, and many people are worried about utility bills. By installing smart home technology running automated security at home, you control anything inside the house while you are anywhere in the world.

Those devices, such as smart security devices that include smart cameras, audio devices, and sensors, make life safer and enable you to check your home at any time you want. What a person needs most when he or she falls into an emergency or a dilemma is someone to help him or her and respond to his or her call, or even communicate his or her voice to the competent authorities, but what if the person is alone in the house when a fire breaks? Our idea is to develop an alarm system for emergencies using Automatic Speech Recognition ASR with Deep Learning, which is a Machine Learning technique that simulates the human neural system in most of its capabilities, including understanding auditory, olfactory, and vision. Because of our specialization in electronics and communication engineering, we can do this project based on data collection and analysis and the development of electronic circuits, such as Arduino microcontrollers. This project will be very beneficial to humanity and will contribute to saving lives.

Some may wonder why we need such a device when mobile phones are common— they can be used for calling for help in emergencies. The answer is that the situation in emergency cases is very difficult and there may be some delay in asking for help, but through this device, you can gain some seconds or minutes to get help as soon as possible, taking into account that this little time will contribute to preserving expensive and densely populated buildings. If we want to implement the project with speech recognition, we have to collect data by taking audio samples with different phrases from both genders and ages, so that these phrases lead to the detection of the disaster according to the type of disaster. For the fair emergency example sound, samples would be taken and labeled whether there is a fire in a place or not, and the number of samples would be enormous. The type of research that we have been working on is quantitative research since deep learning is a data-driven approach. These samples are considered original because they are collected by the members of this project research team. The research tools that are used in this research project depend on collecting audio samples from volunteers and on preparing questionnaires about the requirements that the end-users of this smart alarm would want to have. The device works in places that do not have Internet primarily, but where can you find high-quality speech recognition data containing what are the exact training specifications you need?

The good news is that you have an existing letter recognition of the Arabic language. Accordingly, Plan B involves using speech-to-text APIs, which is a technology for converting spoken words or audio content into text. This can be It is accomplished by existing applications, APIs, tools, and other software solutions. So, speech-to-text APIs are APIs that transcribe audio into written text. It uses machine learning and artificial intelligence trained by Big Data to discover patterns in speech waveforms for the transcription that makes us capable to implement a contingency plan with those services, such as IBM Watson Speech to Text, Google Cloud speech-to-text, Amazon Transcribe, and Microsoft Azure text-to-speech services [1]. There are many problems that our research contributes to solving. Demonstrating these issues will be the primary statement in introducing our new alarm sound system. The following points focus on the most important problems such as wrong location, people busy with daily life, waste of time, and easy communication. And the research aims to solve these problems using: Collecting data and samples, Arduino, microphones, sensors, and deep learning.

1.1. Arduino Nano 33 BLE Sense

Since its launch in 2005, the Arduino Nano 33 BLE Sense platform, shown in Figure 1, has developed into one of the most well-known names in the field of embedded computing. Moreover, it is an open-source electrical platform built on simple hardware and software. As shown in the pinout layout of Figure 2, an LED can be turned on, or something can be published online by using an Arduino board to read inputs like light from a sensor or a user's finger on a button. By giving the board's microcontroller a set of instructions, you can direct your board's actions. Use the processing- and wiring-based Arduino Software IDE and the Arduino programming language to accomplish this. Students, hobbyists, artists, programmers, and professionals from all over the world have flocked around this open-source

platform, and their contributions have added up to an astounding quantity of knowledge that is easily accessible and may be very helpful to both beginners and specialists. At the Ivrea Interaction Design Institute, Arduino was created as a simple tool for quick prototyping geared toward students with no prior experience in electronics or programming. The Arduino board began altering as soon as it gained a larger following, distinguishing its offering from basic 8-bit boards to items for Internet of Things (IoT) applications, wearables, 3D printing, and embedded settings.

1.1.1. Why Arduino Nano 33 BLE Sense?

Arduino has been utilized in countless projects and applications due to its straightforward and user-friendly interface. The Arduino software is simple for novices to use yet flexible enough for experts. It is supported by Mac, Windows, and Linux. It is used by instructors and students to create inexpensive scientific equipment, demonstrate chemistry and physics concepts, or begin learning programming and robotics. Interactive prototypes are created by designers and architects, and musicians and artists utilize them for installations and to test out new musical instruments. Of course, makers use it to construct many of the projects displayed, for instance, at the Maker Faire. A crucial tool for learning new things is Arduino. Anyone can start tinkering by simply following the step-by-step instructions of a kit or exchanging ideas online with other members of the Arduino community, whether they are kids, hobbyists, artists, or programmers. For physical computing, there are numerous additional microcontrollers and microcontroller platforms available. Other apps with comparable features include MIT's Handyboard, Phidgets, Parallax Basic Stamp, and many more. All of these tools take the complex microcontroller programming intricacies and put them in a convenient form. The use of microcontrollers is also made simpler by Arduino, but compared to other platforms, it has the following benefits for educators, students, and curious amateurs according to the cited references:

- Affordable: In comparison to other microcontroller platforms, Arduino boards are reasonably priced. Even the pre-assembled Arduino modules cost less than $70.
- Cross-platform: The Arduino Software (IDE) is compatible with Linux, Macintosh OSX, and Windows. The majority of microcontroller systems are Windows-only [2].
- Anybody looking to start using embedded machine learning, whether they are a beginner, maker, or professional, should consider the Arduino Nano 33 BLE Sense. It is based on the Arm Mbed™ OS and the nRF52840 microcontroller. Easy-to-use, straightforward programming environment: The Arduino Software (IDE) is flexible enough to be useful to both novice and advanced users. It's based on the Processing programming environment, which makes it easier for teachers and ensures that students learning to program in that environment are already familiar with the Arduino IDE [3].
- Open source and extensible software: Arduino software is made available as an open-source tool that seasoned programmers are free to modify. C++ libraries can be used to expand the language, and those interested in technical details can switch from Arduino to the AVR-C programming language, on which it is based. Similarly to that, if you like, you can directly incorporate AVR-C code into your Arduino applications [4].
- Open source and extendable hardware: The Arduino boards' plans are made available under a Creative Commons license, allowing skilled circuit designers to create their own version of the module while modifying and extending it. The breadboard version of the module can be constructed by even relatively inexperienced users in order to comprehend its operation and save money.
- The TinyML and TensorFlow™ Lite capability: Arduino Nano 33 BLE Sense combines a small form factor with a variety of ambient sensors. The Nano 33 BLE Sense will make the process simple, whether you want to build your first embedded ML application or use Bluetooth Low Energy to link your project to your phone.
- Several Sensors: The Nano 33 BLE Sense is equipped with sensors that can detect color, proximity, motion, temperature, humidity, audio, and more in addition to the ability to communicate via Bluetooth Low Energy.

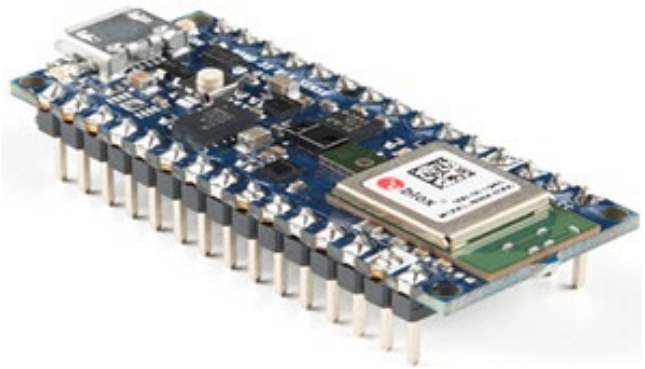

Figure 1: The Arduino Nano 33 BLE Sense (htt1)

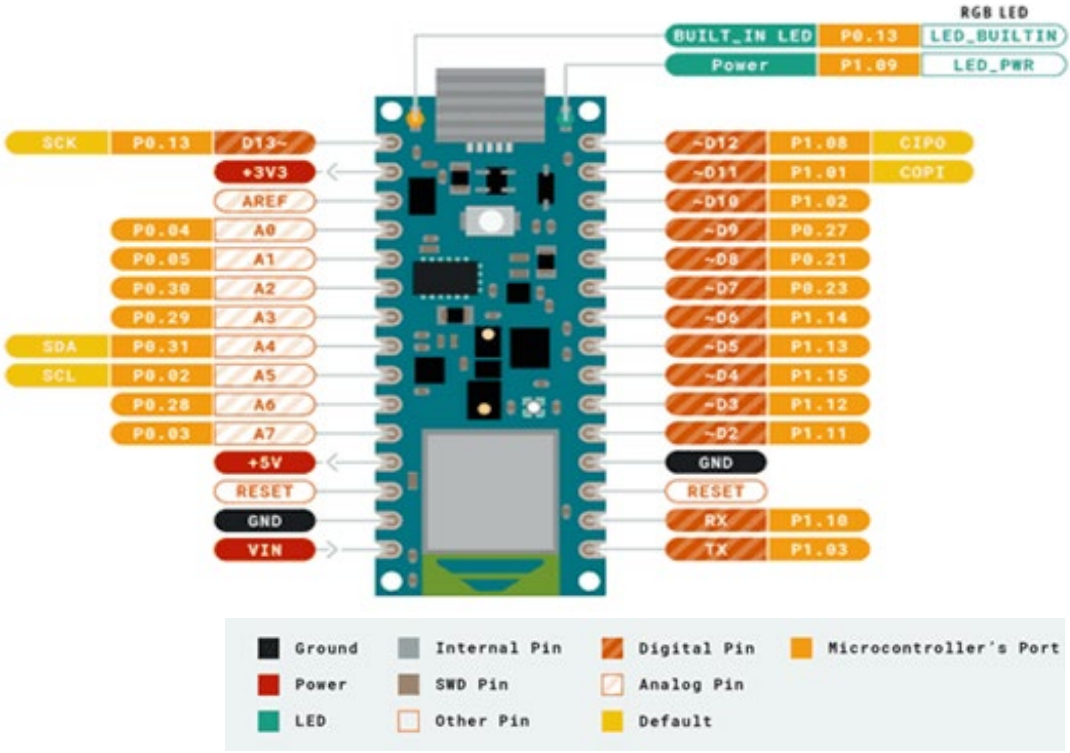

Figure 2: The pinout for Nano 33 BLE Sense [5].

The microcontroller on the Arduino Nano 33 BLE Sense runs at 3.3V, which means that you must never apply more than 3.3V to its Digital and Analog pins, shown in Figure 3. Care must be taken when connecting sensors and actuators to assure that this limit of 3.3V is never exceeded. Connecting higher voltage signals, like the 5V commonly used with the other Arduino boards, will damage the Arduino Nano 33 BLE Sense. In Figure 4, the LSM9DS1 inertial measurement unit features a 3D accelerometer, gyroscope, and magnetometer and allows you to detect orientation, motion, or vibrations. The APDS9960 chip, in Figure 5, allows for measuring digital proximity and ambient light as well as for detecting RGB colors and gestures. Marked in Figure 6, the HTS221 capacitive digital sensor measures relative humidity and temperature. It has a temperature accuracy of ± 0.5 °C (between 15-40 °C) and is thereby perfectly suited to detect ambient temperature. Figure 7 shows the LPS22HB, which picks up on barometric pressure and allows for a 24-bit pressure data output between 260 to 1260 hPa. This data can also be processed to calculate the height above sea level of the current location. Located in Figure 8, the MP34DT05 is a compact, low-power omnidirectional digital MEMS microphone with an IC interface [6,7]. It has a 64 dB signal-to-noise ratio, is capable of sensing acoustic waves, and can operate in temperatures of -40 °C to +85 °C.

The u-blox NINA-B306 module, shown in Figure 9, provides Bluetooth functionality for the Nano 33 BLE Sense. The Arduino BLE library can be used to operate this module. Together with its remarkable array of sensors, this board's key selling point is its ability to run TinyML-based Edge Computing applications (AI). TensorFlow™Lite and the Arduino IDE can be used to build your machine learning models and upload them to your board. Sandeep Mistry, an Arduino developer, and Dominic Pajak, an Arduino adviser, have created an introduction to AI tutorial for the Nano 33 BLE Sense. "Arduino Nano 33 BLE Sense" is its name. The reason it is dubbed "Nano" is that it resembles the original Arduino Nano in terms of size, pinout, and form factor. It is actually intended to be used in place of the original Arduino Nano because the new module runs on 3.3V rather than the original Nano's 5V. In order to denote that the board runs on 3.3V, I believe this is where the name "33" comes from. The name "BLE" then denotes that the module is Bluetooth Low Energy (BLE5 5.0) compatible, and the name "sense" denotes the presence of on-board sensors such as an accelerometer, gyroscope, magnetometer, temperature and humidity sensor, pressure sensor, proximity sensor, color sensor, gesture sensor, and even a built-in microphone. This is how an Arduino Nano 33 BLE sense board appears right out of the box; we shall discuss BLE and other sensors in more detail later. [7]

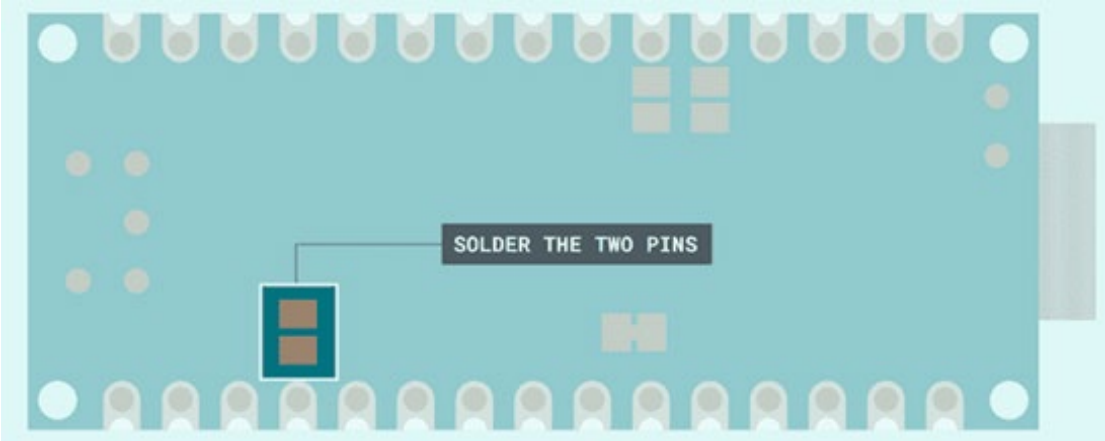

Figure 3: Soldering the VUSB pins.

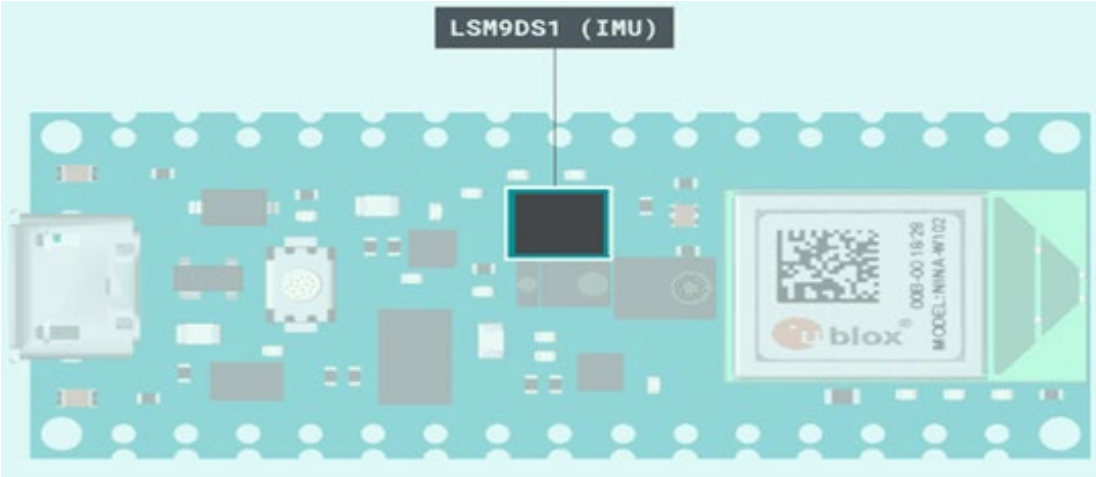

Figure 4: The LSM9DS1 Sensor.

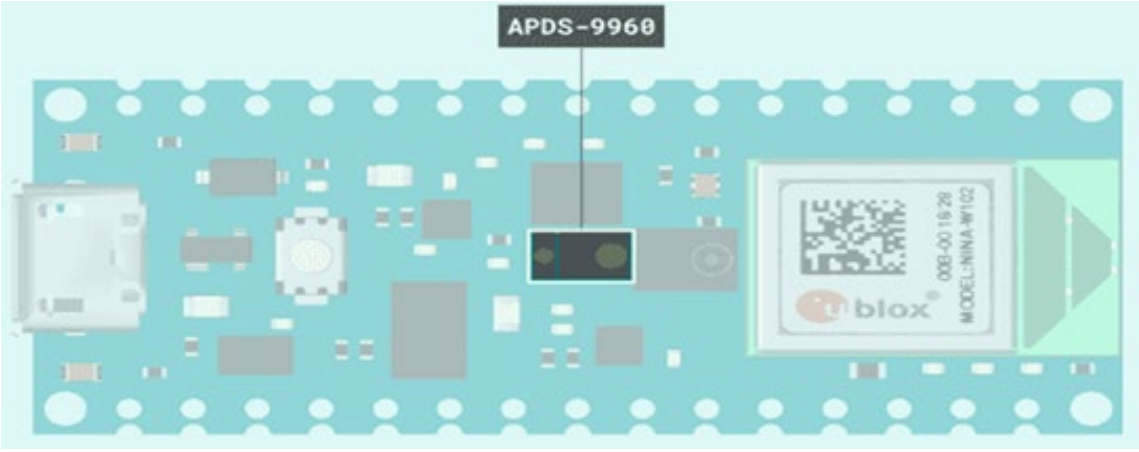

Figure 5: The APDS-9960 Proximity and Gesture Sensor.

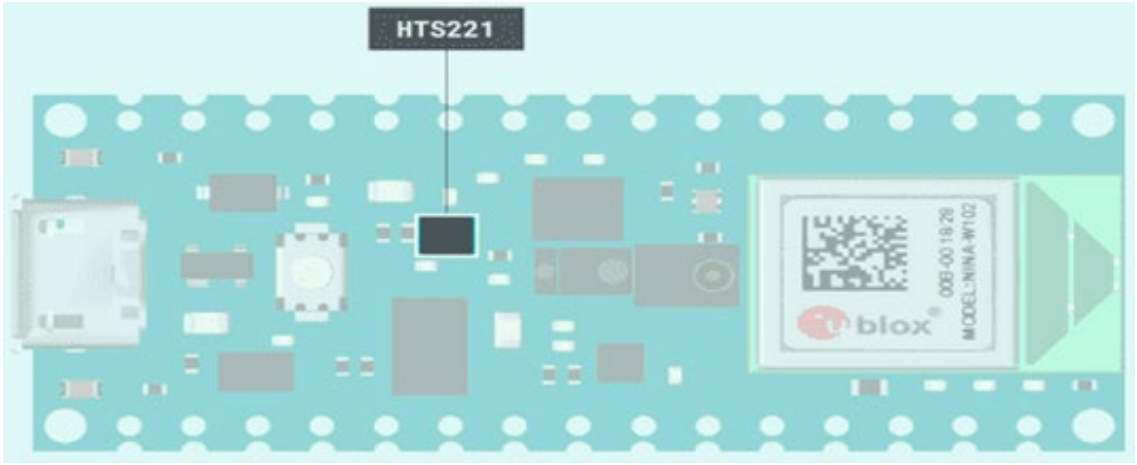

Figure 6: The HTS221 Temperature and Humidity Sensor.

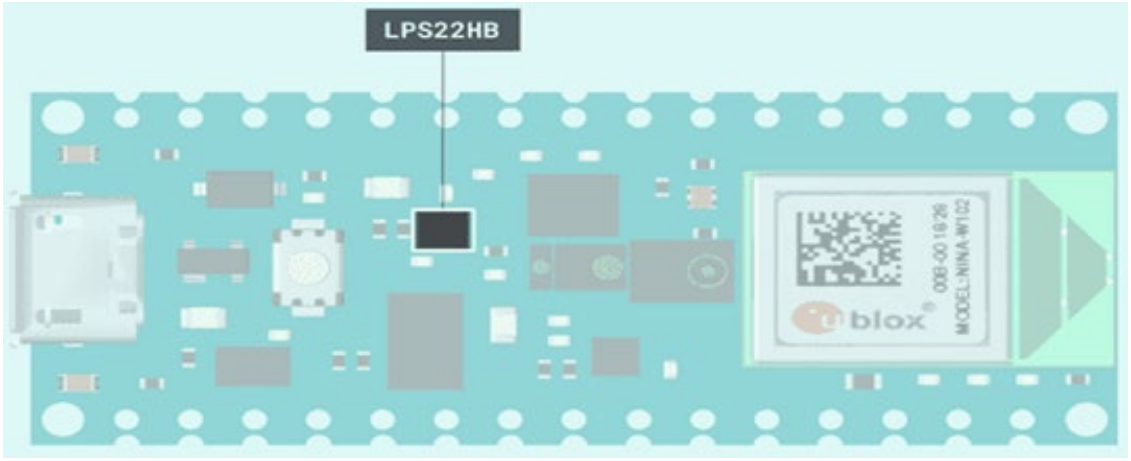

Figure 7: The LPS22HB Pressure Sensor.

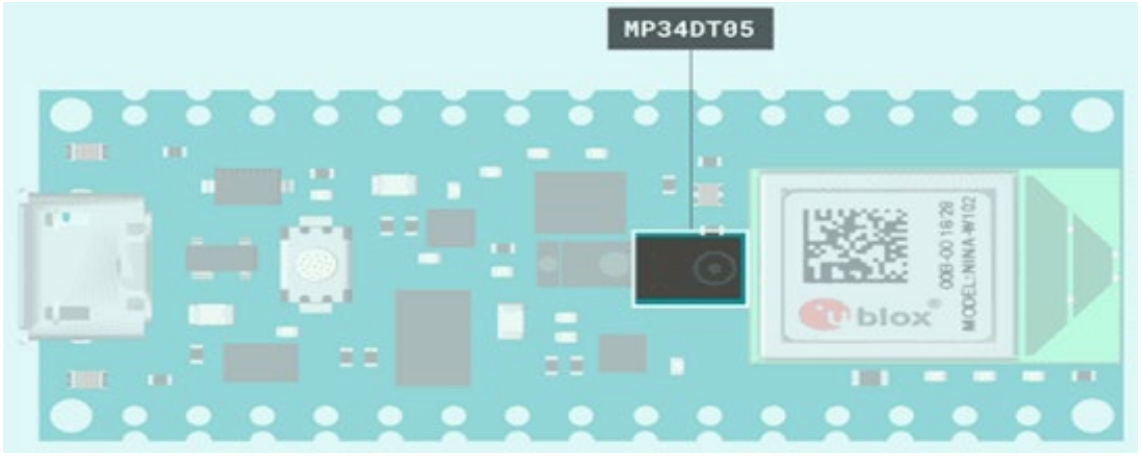

Figure 8: The MP34DT05 Microphone.

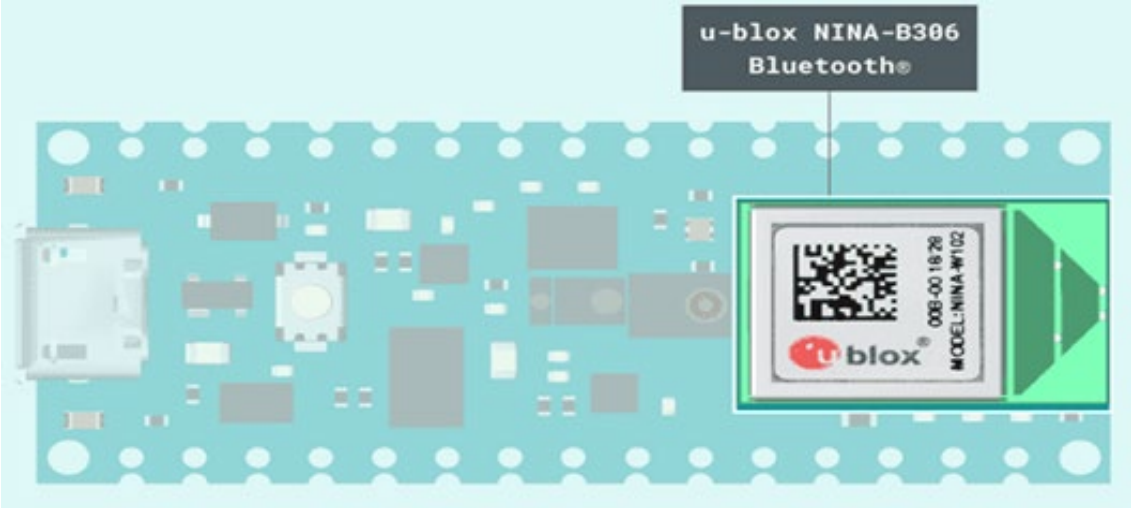

Figure 9: Bluetooth Module.

1.2. Bluetooth

The telecoms company ERICSSON started a study in the year 1.994 to look into the potential of a radio low-cost interface between mobile phones and accessories. The goal was to do rid of the wires that connect mobile phones to PC cards, headphones, desktop gadgets, etc. Beginning in 1997, Ericsson establishes tighter ties with other producers of portable electronics to spur demand for this technology. The idea was straightforward: in order for the system to be effective and truly usable, a significant number of portable devices needed to employ the same technology. A Group of Special Interest was established in February 1998 by five companies: Ericsson, Nokia, IBM, Toshiba, and Intel (SIG). In terms of business, this group has the ideal blend of two market leaders in mobile phones, two market leaders in desktops and laptops, and a market leader in the processing of digital signals technology. The formulation of a worldwide definition for communication without wires of short range was the aim. On May 20 and 21, 1998, the general public in Tokyo, San Jose, and London were informed of Bluetooth's collaboration (Japan). A number of companies embrace the technology as a result of that global statement. The consortium's goal was to create a common device and the software that operates it. It might be fascinating to explain the history of Bluetooth's name and emblem before getting into the technical aspects. Everyone owns a smartphone with the Bluetooth logo, but few people are aware of its significance. [8]

The Bluetooth standard was originally conceived by Dr. Jaarp Haartsen at Ericsson in 1994, more than 20 years ago. It was given that name in honor of King Harald Gormsson, a great Viking and ruler who brought Denmark and Norway together in the tenth century. Dr. Haartsen was given the task of creating a short-range wireless connection standard that could take the role of the 1960s-era RS-232 wired telecommunications standard, which is still in use today. The technology used by Bluetooth® is known as short-link radio. It runs at the unlicensed but regulated, 2.4 to 2.485GHz range and it employs radios to communicate and make connections between two or more devices. The frequency-hopping spread spectrum technique, which is the foundation of Bluetooth, was first introduced in the 1940s by composer George Antheil and actress Hedy Lamarr. Lamarr and Antheil wanted to find a means to guard against radio-guided torpedoes being jammed.

In its most basic form, Bluetooth is a piconet, a type of short-range wireless network. Other businesses than Ericsson that had the idea of a short-range wireless communication between electronic equipment in 1994 included Intel, Nokia, IBM, and Toshiba. These businesses realized at the time that a protocol needed to be defined so that it could be utilized by all electronic devices in order to develop a short-range wireless link. These businesses established the Bluetooth S IG in 1996, and it was formally founded in 1998. SIG had only 5 members when it initially began, but by the end of its first year, it had 4,000, and today it has more than 30,000. In the same way that King Harald Gormson unified the Danish tribes into a single kingdom, Bluetooth aims to unite devices. About 1999 saw the debut of Bluetooth 1.0, followed by Bluetooth 2.0 in 2004, Bluetooth 2.1 in 2007, Bluetooth 3.0 in 2009, Bluetooth 4.0 in 2010, Bluetooth 4.1 in 2013, Bluetooth 4.2 in 2014, Bluetooth 5.0 in 2016, and Bluetooth 5.1 in 2019.

Three operating modes (BR, EDR, and HS) are listed in the Bluetooth® 3.0 specification, which you may obtain online (AMP) For simplicity, users typically refer to these three operating modes as classic Bluetooth®. Wibree, a wireless technology created by Nokia, Nordic Semiconductor, and other businesses with the aim of discovering a low-power wireless communication technology for electrical devices, and SIG combined in 2010. Wibree was renamed Bluetooth Low Energy by SIG. In order to dramatically minimize power usage, Bluetooth Low Energy was created to cut down on the amount of time the Bluetooth radio is turned on. From the Bluetooth 4.0 specification, both Traditional Bluetooth® and Bluetooth Low Energy have been incorporated; nevertheless, they operate differently and are incompatible with one another. Several physical layer modulation and demodulation techniques are used for mode, traditional Bluetooth, and Bluetooth Low Energy. As a result, Bluetooth Low Energy and standard Bluetooth are incompatible. In general, Bluetooth classic is mostly utilized for audio applications (wireless headphones, for instance), whereas Bluetooth Low Energy is more frequently employed in applications that require less power (such as wearables and IoT devices, for example).

1.2.1. How Does Bluetooth Low Energy Work?

To understand how Bluetooth Low Energy works, we need to talk about the roles and responsibilities of two devices that are connected through Bluetooth. In any Bluetooth connection, as in Figure 10, two roles are being played: the central and peripheral roles. Devices with a central role are also called servers while devices with a peripheral role are also called clients.

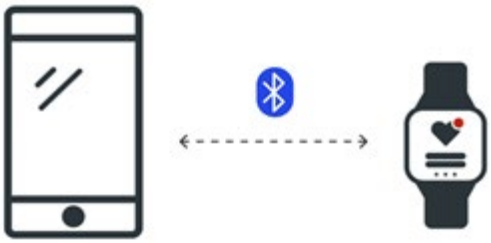

Figure 10: Central and Peripheral Roles in Bluetooth Applications [8].

One device, the peripheral, will broadcast or promote information about itself to any nearby devices once a Bluetooth connection has been made. Another device, the central, will be conducting a scan at the same time and listening for any device(s) that are broadcasting data. An attempt will be made to connect the peripheral device as soon as the central device receives the advertising data from it. Once a connection has been made, the central device will communicate with the peripheral device's information once it is accessible. This information exchange is accomplished through the use of services. A collection of capabilities is a service. For instance, a smartwatch can monitor your heart rate, log your daily exercise, and monitor your sleeping patterns. For instance, a service called "health service" might include these three features. Central devices enable peripheral devices to quickly identify, choose, and engage with the required services they want by grouping capabilities in services. Every service has a special identifying number, or UUID. For official Bluetooth® specification services, this code can be 16-bit or 32-bit long, however non-official Bluetooth® specification services (the ones we can design) are 128-bit long and UUIDs can be generated at random. A combination of services make up a profile. There will be a list of attributes for each service. The center device's abilities are each represented by one of these attributes. The health service in the prior scenario would have three qualities (heart rate, physical activity, and sleep pattern). The peripheral device can write information to, request information from, and subscribe to updates from these characteristics after it has discovered them. Every characteristic has a 16-bit or 128-bit long UUID, much as the services.

The Nano 33 BLE Sense, the primary device, will establish a connection with the Nano 33 BLE, the peripheral device, and search for the gesture Service. If the central device uses its gesture sensor to detect a gesture after the connection between the central and peripheral devices has been made, it will record the type of gesture it has detected in the gesture type characteristic of the gesture Service, as depicted in Figure 11, an. Following that, the built-in RGB LED of the peripheral device, the Nano 33 BLE, will switch on a certain hue based on the value contained in the gesture type characteristic [9].

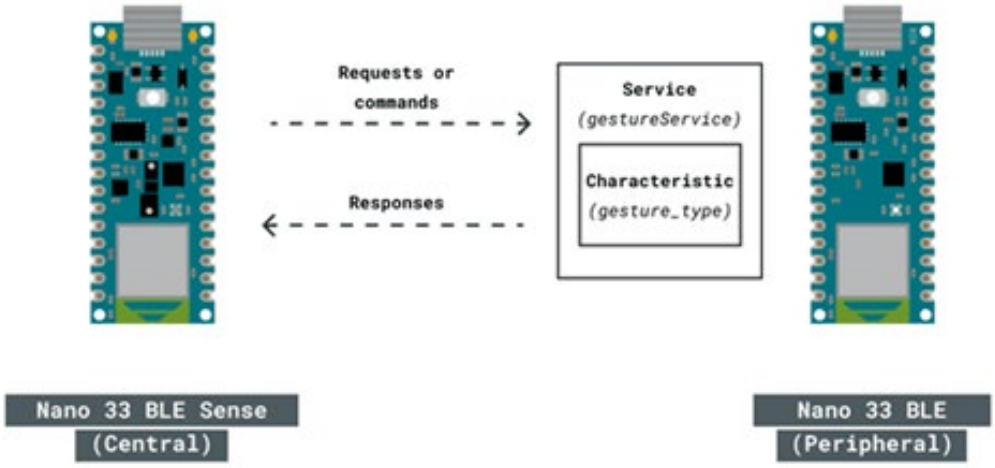

Figure 11: Gesture Example Architecture.

1.3. Deep Learning

Due to its potential advantages, such as feature extraction and data classification issues, deep learning has attracted growing attention in a variety of fields, including automatic speech recognition, computer vision, natural language processing, drug discovery toxicology, audio recognition, bioinformatics, and automatic driving of vehicles. The entire potential cost benefits for maintenance and renovation activities are increasing due to the dynamic nature of the research field and its many applications. Machine learning techniques help in modeling high-level abstract representations of data by using processing layers that include complicated structures. Deep learning is a widely used technology. With a vast amount of unlabeled data, the software tools in this field offer finer representations. Deep learning

software recognizes patterns in digital representations like data, music, graphics, etc. Deep Learning has been at a "Permanent Peak" since 2015, according to Gartner's hype cycle, and according to a Higher Frequency Subband Research survey, 86% of respondents think the technology has a significant economic influence on their industry. It is a fast expanding field with a variety of potent approaches and enormous computing power, where the computer recognizes objects and instantly interprets spoken language. Models with several non-linear information processing stages or layers and approaches for supervised or unsupervised learning for feature extraction at an abstraction level are the two primary components of deep learning. The size of the training data set, chip processing power, and current developments in signal processing and machine learning research are key factors in its appeal. In order to effectively use both labeled and unlabeled data, deep learning approaches are effectively leveraging complex nonlinear functions to acquire hierarchical and distributed feature representations. The various deep learning approaches and structures, like convolutional Deep Neural Systems Continuous Neural Networks Long short-term memory, artificial neural networks, and deep belief networks are discussed.

[https://www.researchgate.net/publication/353439278_Introduction_to_Deep_Learning]

Moreover, a more advanced subset of machine learning known as "Deep Learning" employs multiple layers of algorithms to analyze data, simulate the thinking process, or create abstractions. It is frequently used to comprehend spoken language and distinguish objects visually. Each layer transmits information to the one underneath it, with the output of one layer serving as the input for another. The input layer is the first layer in a network, and the output layer is the last. Hidden layers are all the layers that exist between input and output. Typically, each layer is a straightforward algorithm with a single type of activation function. Deep learning also includes feature extraction. It is employed in image processing and pattern recognition. To automatically create useful "features" of the data for training, learning, and comprehension, feature extraction employs an algorithm. A data scientist or programmer is often in charge of feature extraction. When Walter Pitts and Warren McCulloch developed a computer model based on the neural networks of the human brain in 1943, deep learning had its beginnings. They imitated the mental process using a set of mathematical formulas and algorithms they named "threshold logic." Since then, there have only been two important interruptions in the development of deep learning. Both have anything to do with the infamous AI winters.

1.3.1. Preparing Audio Data for a Deep Learning Model

Machine learning applications of computer vision previously relied on conventional image processing methods to do feature engineering in the years prior to Deep Learning. For instance, by employing algorithms to find corners, edges, and faces, we would produce hand-crafted features. We would rely on methods like extracting N-grams and determining term frequency for applications that use natural language processing. Similar to this, traditional digital signal processing methods were used by audio machine learning programs to extract features. For instance, phonetics principles could be used to analyze audio data and extract components like phonemes in order to understand human speech. To tackle these issues and optimize the system's performance, a lot of domain-specific knowledge was needed. Yet, as Deep Learning has spread more widely in recent years, it has also had great success in processing audio. With deep learning, conventional audio processing methods are no longer necessary, and we can rely on common data preparation without needing to generate many manually-created features. The fact that we don't actually work with audio data in its raw form when using deep learning is more intriguing. Instead, a typical method is to transform the audio data into images before processing those images with a conventional Convolutional Neural Network (CNN) architecture! Really? Transform audio into visuals? That has a science-fiction feel about it. Of course, the response is really banal and ordinary. By creating Spectrograms from the audio, this is accomplished.

1.4. Sensors

A sensor is a piece of equipment that receives a signal or stimulus and reacts to it by producing an electrical signal, illustrated in Figure 12. A few types of electrical signals, such as current or voltage, are represented by the output signals. The sensor is an apparatus that collects various signals, such as physical, chemical, or biological signs, and transforms them into an electric signal. Based on the applications, input signal, conversion method, material utilized, and sensor properties like price, accuracy, or range, the sensors are divided into many types. The classes of sensors in this chapter include thermal, magnetic, optical, mechanical, and chemical. Together with an introduction to the fundamental types of sensors, the transfer functions, properties, and specifications are also covered.

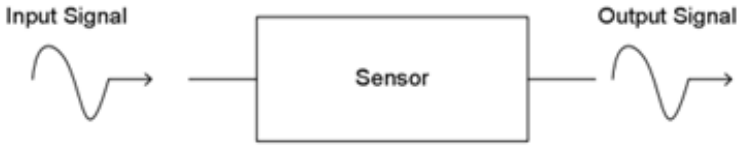

Figure 12: A Drawing that Describes the Sensor's Work

And it works to make our lives easier in a way that was not possible before, as the sensors can perform a variety of functions while you are at your location and without the need for your help, such as turning on the lights, adjusting the room's temperature, detecting the presence of a fire, opening the garage door or the mall's door, among other things. A sensor is a device or component that aids in detecting any changes in a physical quantity, such as pressure or force, or an electrical quantity, such as current or any other form of energy. Following the observation of the changes, the sensor sends the detected input to a microcontroller or

microprocessor, and based on that input, the microcontroller or microprocessor makes a decision. In response to a change in the input signal, the sensor generates a readable output signal that may be optical, electrical, or take any other form. Sensors are crucial to any measurement system. In actuality, the sensors are listed first in a system's block diagram. Direct interaction with the variables through measurement results in a comprehensible output. For instance, the autonomous flight control system uses a number of sensors to perform a variety of functions, including fuel level monitoring, maneuvering, location tracking, door status, and obstacle detection, cruise control, and altitude control.

All of this sensor data is processed by a computer by comparing it to preset values. The computer then sends command signals to numerous components that help with smooth flying, such as the engines, flaps, rudders, actuators, etc. The airplane can operate on autopilot thanks to a mix of sensors, computers, and mechanics. Depending on its features, a device can be categorized as a sensor. These features include:

• An object that functions as an energy transducer and has electrical characteristics that influence energy transfer.
• A tool that keeps track of alterations to the physical environment.
• Possess a transformer for power input.
• The sensor is an electronic system component.

Moreover, many writers and professionals have created a number of classes of sensors. Some are relatively straightforward, while others are intricate. An expert in the field may already use the classification of sensors that follows, although it is fairly straightforward. The sensors are split into Active and Passive categories in the initial classification. Active sensors are those that need an external power signal or an excitation signal. Conversely, passive sensors directly produce output response without requiring any external power supply. The other classification method is based on the sensor's method of detection. Electric, biological, chemical, radioactive, and other methods of detection are some examples. The following classification is based on the input and output of a conversion, or event. Photoelectric, thermoelectric, electrochemical, and electromagnetic, thematic, etc. are a few of the frequently occurring conversion phenomena. Analog and digital sensors make up the last two categories for the sensors. In relation to the amount being measured, analog sensors produce an analog output, or a continuous output signal (often voltage but occasionally other variables like Resistance, etc.).

In contrast to analog sensors, digital sensors operate on discrete or digital data. Digital sensors, which are utilized for conversion and transmission, only store and transmit digital data. The list of numerous sensor types that are frequently used in various applications may be seen below. The purpose of each of these sensors is to measure a certain physical characteristic, such as temperature, resistance, capacitance, conduction, heat transfer, etc.

- Sensor for Temperature
- Nearby Sensor
- Accelerometer
- The IR sensor (Infrared Sensor)
- Sensor for Pressure
- Light Detector
- Sensor using Ultrasound
- Gas and Smoke Sensor

Medical professionals can also use sensors, such as the mercury thermometers created in 1714 by Daniel Gabriel Fahrenheit. The scale consists of a glass cylinder with a mercury tank at one end. The amount of mercury in the cylinder can be used to determine the temperature. The concept behind the scale is based on how mercury expands at high temperatures and contracts at cold temperatures. The remainder of the cylinder is either drained of air or filled with nitrogen. Anders Celsius developed a thermometer scale based on the ice-melting and water-boiling points in 1742. The gradation was initially reversed because ice melts at zero degrees Celsius while water boils at 100 degrees Celsius (212 degrees Fahrenheit) (32 degrees Fahrenheit). A year later, Jean-Pierre Christen, a Frenchman, modified the Salesian system and gave it the name Celsius system. To make reading easier, certain thermometers have the capacity to maintain the top reading for a longer amount of time. This is accomplished by lowering the size of the cylinder that is attached to the tank in order to make it more difficult for mercury to return to the tank. Via the use of a medical thermometer, these scales are used to assess body temperature.

For our project, we employed the MP34DT05 microphone sensor, a crucial sensor because it transforms acoustic signals into digital information. Mobile terminals, voice recognition programs, and even gaming and virtual reality input devices frequently use microphones. The MP34DT05-A (Figure 13) is a digital MEMS microphone that is extremely small, low-power, omnidirectional, and designed with a capacitive sensing element and an IC interface. The sensing component, which can detect acoustic waves, is created utilizing a unique silicon micromachining procedure intended for the creation of audio sensors. Two sensors can be synchronized as in Figure 14. Using a CMOS manufacturing method, the IC interface can be designed with a specialized circuit that can deliver a digital signal externally in PDM format. Also, the MP34DT05-A is a low-distortion digital microphone with a 64 dB signal-to-noise ratio and a sensitivity of -26 dBFS 3 dB, as was already indicated. The MP34DT05-A is guaranteed to function throughout a wide temperature range, from -40 °C to +85 °C, and is offered in a top-port, SMD-compliant, EMI-shielded package.

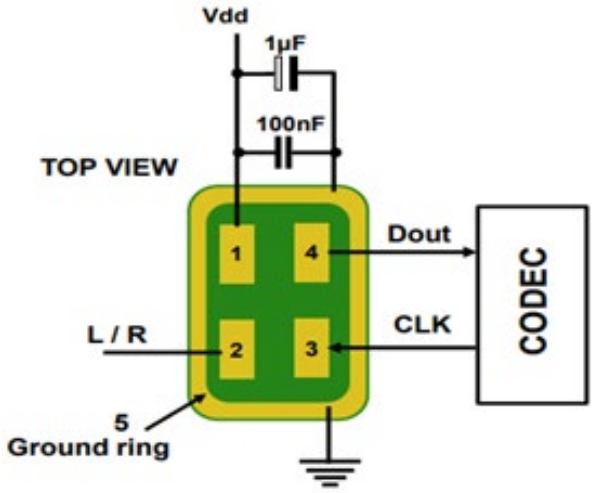

Figure 13: The MP34DT05-A Electrical Connections for Mono Configuration

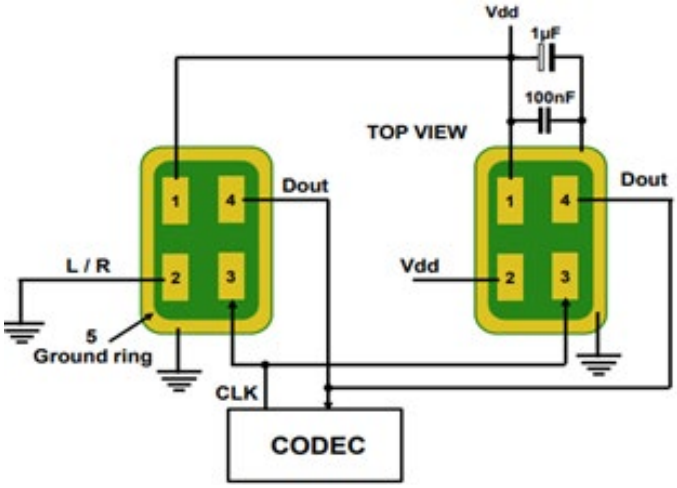

Figure 14: The MP34DT05-A Electrical Connections for Stereo Configuration

There are many sensors that can be used in our project, such as: Buzzer sensor (Figure 15) is a simple yet effective part that gives our project or system sound features. Its 2-pin structure is very small and compact, making it simple to utilize on a breadboard, perf board, or even a PCB, making it a common component in most electrical applications. Here is a straightforward buzzer that, when energized, emits a constant beeeeeppp sound. Just providing this buzzer with a DC power supply that ranges from 4V to 9V will enable you to utilize it. A basic 9V battery can also be utilized, however a regulated +5V or +6V DC supply is advised. The buzzer is typically connected to a switching circuit that allows it to be turned ON or OFF at the appropriate time and interval.

Figure 15: A 5-volt Buzzer Sensor

Thermistor sensor (Figure 16), whose resistance fluctuates as the temperature changes are temperature-dependent resistors. They are extremely sensitive and respond to even the smallest temperature changes. They work well when temperatures must be kept within 50°C of ambient or when a specified temperature must be maintained. And a range of ordinary items depend on temperature sensors. For instance, maintaining and controlling temperature is essential for the correct operation of domestic appliances like ovens, refrigerators, and thermostats.

Figure 16: A 10K Thermistor Sensor

Flame sensor (Figure 17): The flame sensor's reaction is typically employed for flame alarm applications because it is the most sensitive to common light. This module is capable of detecting flames and light sources with wavelengths between 760 and 1100 nanometers. The IO port of the microcomputer can be directly coupled with a single chip and a small plate output interface. To prevent damage from high temperatures, the sensor and flame should maintain a specific distance from one another. If the flame is larger, test it at a greater distance. The shortest test distance is 80 cm. The flame spectrum is highly sensitive because the detecting angle is 60 degrees. The flame spectrum is very sensitive due to the 60 degree detection angle. If there is a fire, it can be put to use in our project.

Figure 17: A Flame Sensor

Gas sensor (Figure 18): Chemical sensors that are of utmost significance are gas sensors. A chemical sensor is made up of a transducer and an active layer that transform chemical data into another type of electrical signal, such as a change in frequency, current, or voltage. Since the environment we live in is primarily populated by humans, plants, and animals, the safety of their lives is of the utmost importance. As a result, it becomes extremely imperative to detect the presence of these gases because they may be hazardous to human health, atmospheric pollutants, or significant to an industrial or medical process.

Figure 18: A Gas Sensor

In general, conventional detection methods are not very reliable since precise real-time measurements of a target gas's concentration are needed to create systems that sound an auditory alarm to alert people when there is a dangerous or poisonous gas leakage. Nonetheless, several gas sensor technologies, such as semiconductor gas sensors, catalytic gas sensors, and electrochemical gas sensors, have been utilized for various types of gas detection for many centuries. It can be applied to our project to find smoke from fires. Additionally, there are numerous significant sensors that could be used in this project and developed further by adding many of them, and this project could be developed later to be more prepared in emergency situations to save people. However, for the time being, we were happy to settle for adding just one sensor, the MP34DT05-A sensor in Figure 13, due to the project's limited time frame.

2. Related Work

Identifying fires is a general and necessary topic in our lives. Methods and means for identifying and detecting fires may differ, but the goal is the same. In our research topic, we provide

a solution to this problem by recognizing the human voice as "the distress call" when fires occur, and according to previous research, identifying fires was through sensors such as heat and smoke sensors, which in turn determine only if the temperatures are high and determine whether there is smoke.

2.1. Fire Detection

In a previous study the researcher developed a fire detection system based on simultaneous measurements of carbon monoxide, gas carbon dioxide (CO2), and smoke. He combined the rates of smoke rise and the concentration of carbon dioxide or carbon dioxide to increase the reliability of smoke detectors for aircraft and reduce the alarm time, like a sensor that works on its own. Based on the results of the researcher, there may be raw materials containing these gases, which in turn will be a source of attention to the device, which may lead to false alarms. Anyway, I see that detecting fires by adding the voice recognition feature to the alarm device can prevent fires in terms of prompt attention and extinguishing the fire. Another research referred to the technique of detecting forest fires using wireless sensor networks and neural networks, to replace satellite-based monitoring, which in turn takes a long survey period and low accuracy [10]. The researcher indicated that the research aims to use wireless sensor networks for forest fires and predict them quickly and accurately to reduce massive fires that occur in forests and reduce their severity by deploying a large number of sensor nodes in the forest and taking measured data such as temperature, relative humidity, and the neural networks take the data as input to produce a weather index that measures the probability of the weather causing the fire. In light of the past, we believe that the use of neural networks is very important nowadays, so we have adopted neural networks to process data "speech recognition" in our topic. We will go into more detail about neural networks in other sections.

2.2. Speech Recognition Using Neural Networks

A previous study was conducted on voice recognition using neural networks to identify people by vital traits to reduce illegal immigration and crime [11]. The researcher identified the character of voice because the bandwidth associated with speech is much smaller than other image-based human traits. This means faster processing and smaller storage space. Based on the points and criteria identified by the researcher in his topic, we agreed with him in terms of the use of neural networks in general, but we disagreed with him in terms of content. While we are in our research topic, our goal is to recognize and identify the word by the device. According to one of the studies the goal of the researcher was accurate recognition of emotions [12]. To elicit an emotional response from robots, computers, and other smart devices, the researcher classified emotions based on the sound criteria for six feelings (happiness, fear, sadness, disgust, anger, and surprise). Our topic agreed with the recognition of feelings for an important reason, which is to determine the accuracy of the alarm when recognizing words. When the word that is to be identified by the device is said in a state of fear, it is a real warning. But if the word was said in the state of happiness, it would be a false alarm, and therefore it was important to take into account this study and consider its results.

2.3. Detect Fires Using Voice Recognition

In an article Shawn Heymel posted on his GitHub page the author outlines a collection of tools and demo projects for creating embedded keyword identification using machine learning, which is, as we mentioned earlier, keyword discovery [13]. The publisher has used "spotting-keywords" audio. And he used the edge impulse website to teach the machine in terms of collecting samples, adjusting the characteristics of the MFCC, and training the machine on neural networks, and this is the real nerve of our topic. This article is very useful to us in terms of research methodology and tools from websites or even software codes. We have been forced to create different benchmarks for their pricing in different parameters and different data. In the previous manuscript, by looking at the previous studies, we discovered what suits us in general and what suits us in terms of methods and tools. Add to that the ability to combine more than one study and filter it to get what we need without an increase or decrease.

3. Methodology

In this chapter, we will talk about several axes related to the methodology. This chapter concludes its importance in that it talks about the type and how the data collected by the team was selected and the method used by the team in data collection, after that the analysis that was carried out to the data team. We believe that the reader will be able to understand and apply the method we have followed and analyze the collected data. It is necessary for a researcher to design a research methodology for the problem chosen. One should note that even if the research method considered for two problems are the same, the research methodology may be different. “It is important for the researcher to know not only the research methods necessary for the research undertaken but also the methodology” [14]. We had to choose an easy-to-apply and useful methodology for the project as much as possible, so that whoever wants to develop on our project will be able to develop quickly and with great effectiveness. We are targeting a large segment of people who suffer the most from these emergency incidents, so we try as much as possible to have our project with a flexible methodology that can be developed by other researchers. At the beginning of the project, we looked at many projects like ours in order to develop our own methodology, through which we collect and analyze data. There were many methodologies, and we decided to follow one of them and develop it as well. In the next lines, we will list this methodology in details.

3.1. The Targeted Data Collected

In the methodology of data collection, we have targeted many segments of society of different ages and different regions as possible, because this diversity and difference in data help to raise the quality of the project and contributes directly to improving the outputs of the project. As for the words and sentences that we asked the participants to say, they are carefully selected because they largely simulate the reality that we live in, in emergency situations and also in natural situations. We have tried as much as possible to consult and investigate about the samples that we want the project to take a reaction when the device hears it. Some members of the evaluation committee gave us some recommendations that should be considered during the data collection, and we followed their advice. The project could have been much broader, but we decided to limit the scope of the

project to the cases of fire, as was recommended. For example, we were able to make the project work in emergency cases such as cases of theft or harassment. That is, if we add words of help in the collected data related to these cases, the project will encompass a wider range of work. As for talking about the number of samples that we collected, we have collected approximately 900 samples from several people of different ages. We also planned to take ready samples from databases prepared in advance by other researchers to increase the number of data in a short time, but we did not find anything like this especially in Arabic. We believe that we have built an initial database for recognizing emergencies by voice, and this work is a foundation that anyone after us can develop. We aspired to collect much more than this number, but we faced some problems related to the availability of time, because as everyone knows, the duration of the semester is 12 weeks, and this is a short time if we want to expand data collection, so we decided to suffice with 900 samples, and it is a very good number. We will talk in detail about sampling information in another section of this chapter. We know very well that when we talk about areas such as deep learning, it means the volume of data matters; the more data we collect, the higher the quality and effectiveness. It is a direct relationship.

3.2. Data Collection Method

To take samples, we needed some steps that we had to follow so that we could collect samples in an efficient and high-quality manner, so we could then analyze and classify them in the Edge impulse platform. We will list these steps in the next lines.

3.2.1. Create an Account on the Edge Impulse Platform

In the data collection methodology, we have decided to do it by Edge Impulse, and Edge Impulse is ushering in the future of embedded machine learning by empowering developers to create and optimize solutions with real-world data [15].

With millions of developers transforming billions of devices into smarter ones, we are making the process of developing, deploying, and scaling embedded ML applications easier and faster than ever. Zach Shelby and Jan Jong boom founded our company in 2019, and we're on a quest to make it easier for developers to make the upcoming wave of intelligent products. We are committed to supporting machine learning applications for good because we think they can benefit society. [16]

3.2.2. Preparing and Downloading the Required Software

In this step, there were three options to complete the rest of the remaining data collection methodology steps. In the first two options, we must download Arduino IDE.

First Option: In this option, we record samples directly from the Edge impulse platform, and this is relatively easy, but the difference between this option and the other options is that there is some difficulty in downloading and installing some programs regarding this first option, and it may take a lot of time. In this option we have followed this site, where we download Node.js and Python 3.

https://docs.edgeimpulse.com/docs/development-platforms/officially-supported-mcu-targets/arduino-nano-33-ble-sense

After that, we the development the board, which is “Arduino nano 33 BLE Sense” and does not come with the right firmware yet. To update the firmware, we needed to download the last version of Edge Impulse firmware and to open the flash script for your operating system. In our case, it was “flash_windows.bat”. Once it appears in the “Your devices” list of the Edge Impulse platform, as shown in Figure 19, you become able to record the samples from Edge Impulse directly by the Arduino device. In the “Data acquisition” page, you will see something like Figure 20. After that, you can choose how long you want the audio to be recorded from “Sample length”, then you record the audio by “Start sampling”. The maximum allowed period is 19 seconds.

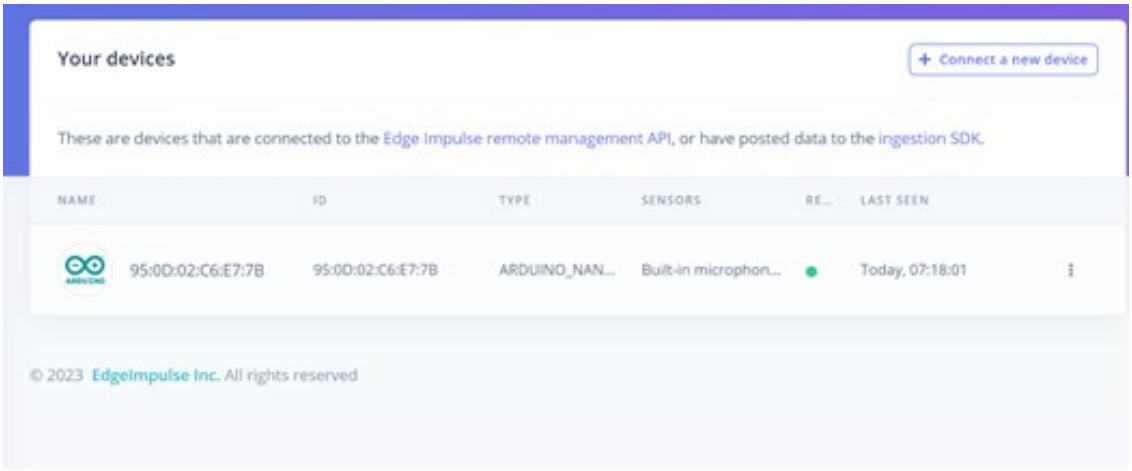

Figure 19: Arduino Microcontroller Connected with Edge Impulse Platform

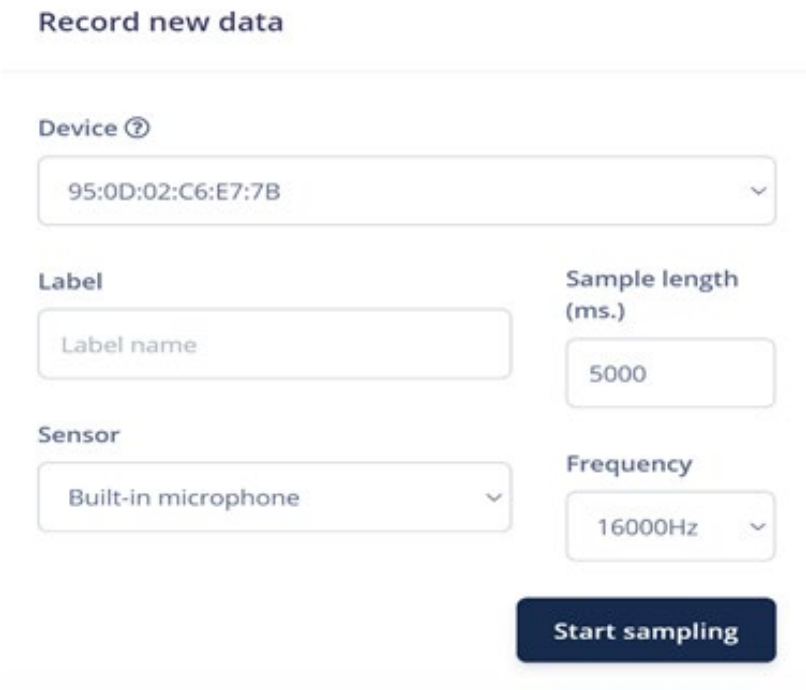

Figure 20: Sampling in Edge Impulse

Second Option: In this option, we need to download another software which is called PuTTY, then we open an Arduino example from the main menu: File > Examples > PDM. Upload the example to compile it on the Arduino 33 BLE Sense board, then follow the steps:

Step 1: Copy the port id of the connected device. To find the connected devices using the Arduino IDE, click Tools > Port.
Step 2: Close the application that uses the port.
Step 3: Redirect the serial communication to text file by PuTTY.
Step 4: When finished, hit the keyboard keys: Control and C.

We will notice here, after recording any sound, that the file will be printed in a text file as shown in Figure 21. Now, we will use Python to convert the file to the Binary format so that we can upload it to the Edge Impulse platform.

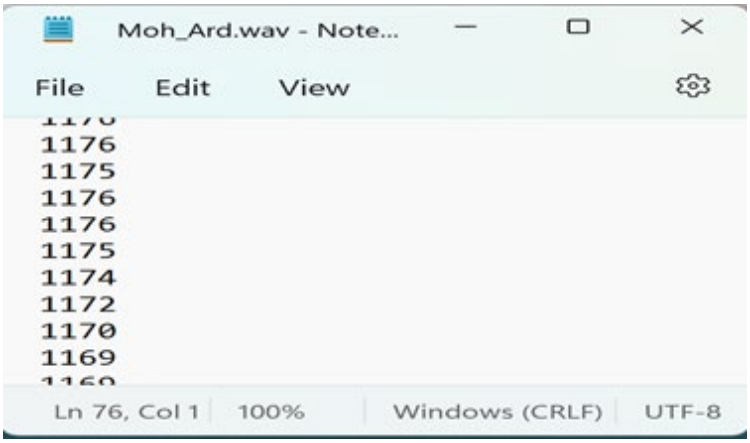

Figure 21: A Text file

Third Option: In this last option, we need the Audacity program only as we use the phone's microphone to record samples. The samples are then uploaded to the computer. Note that the recording that comes from the phone is in MP3 or OGG format, and it also comes with a Sample rate not equal to 16 kHz. This makes the audio file need further processing for the project. We need to convert it to the WAV format and to change the Sample rate to 16 kHz. To do so: open the file from Audacity, change the Sample rate to 16 KHz, and select Tracks > Resample, as shown in Figure 22. Finally, change the file format when saving the file: Export > Export as WAV. This page appears in Figure 23.

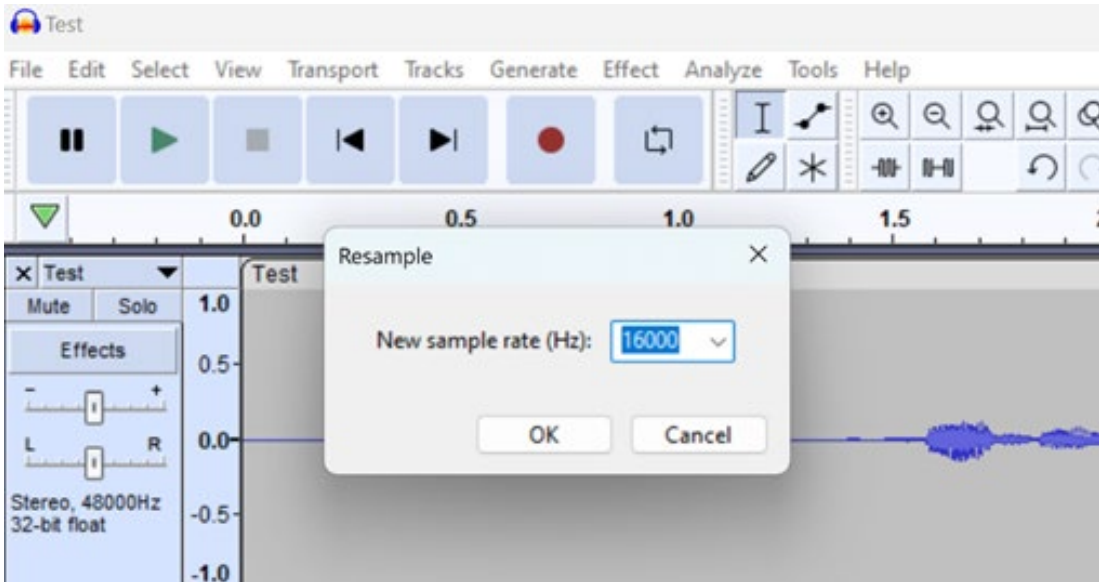

Figure 22: Resampling Process

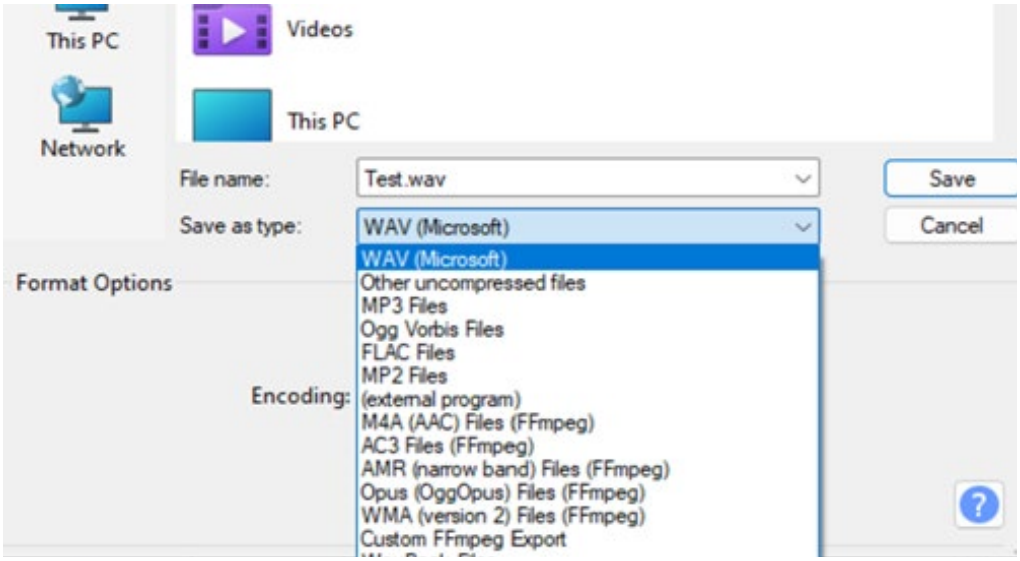

Figure 23: Saving the file as WAV

Make sure to save the file in WAV (Microsoft). Then, the audio sample that we recorded from the mobile phone is ready to be uploaded to the Edge Impulse platform, It is worth noting that we prefer to use the first two options, and the reason for that is that if we think in the long term, we will use Arduino Nano 33 BLE Sense as a receiver for audio signals to respond to them, so it is better to use the same device in the data collection process, because we want to make the entered data and the received data as close as possible. If we use the mobile microphone as a data collector, a slight difference between the input data and the received data may affect the classification performance since the mobiles the Arduino Nano 33 BLE Sense have different microphones' characteristics. Some team members had difficulty and some problems in using the first two options, which led them using the third option. In short, it was valuable that all data be collected by the first two options.

3.2.3. Preparing Consent Form to Participate in Research

This step is one of the most important steps for data collection, because this form contributes to understanding and preserving everyone's rights. We have tried as much as possible to make this form as professional as possible, so that it ensures the participants that we are doing serious work. And it included some themes, namely:

• The title of the research, and the group that is doing the research, as well as the party to which the researchers belong.
• The name and mobile number of the researcher who collects the data.
• It was mentioned that this work done by the participant is voluntary and he or she has the right to refrain, and that he or she can withdraw the samples that were issued from him or her at any time.
• The participant has the right to ask and inquire about anything related to the project.
• Mention the purpose of the project and that this project, if successful, will contribute to helping others.
• The participant identify will be completely protected and confidential.
• The written sentences for the two natural and emergency cases

that the participant will say.
• Among the benefits that the participant will gain is obtaining a certificate of thanks after completing the project, obtaining experience in areas that may be considered new to the participant, spreading the culture of participation in research.
• At the end, a place for the participant's signature and his or her name.
• We must mention here that we have prepared two forms: for children and for adults, and they contain the same axes, and the only difference is that in the children's form, we ask them to have a supervising guardian, and the consent is from the guardians and not the children themselves. These forms can be found in the APPENDIX of this report.

3.2.4. Preparing a Video Clip for the Participants

The clip consists of several parts, and we will list its details in the coming lines.
• The title of the research and that this research studies the relationship between the characteristics of the tone of voice and feelings of fear to activate the automated response to distress calls when a fire occurs in residential places.
• In this part, we want to make the participants prepare to say some words in a real tone, so we ask them to say some words until they do the expected action.
• Showing a countdown timer makes the participant able to get ready to participate.
• Presenting the words that the participant will say, along with showing a video clip of a person sitting in a room where the fire started. The aim of this part is to help the participants act fearful so the word may be uttered in genuine tones that simulate real fire emergency situations.
• We do the same work, but with the other case, which is the normal case, where we try to show the participants a video clip that makes them feel safe and relaxed, so that they say the words in a way that is close to the normal case.
• In the final part, we thank the participant for helping us.

3.2.5. Taking the Samples from the Participants

After preparing the form and the video, and preparing and downloading the required software, we can take samples from the participants. All that remains at this stage is that we go to the adult participants or children and ask them to read the form and then sign in case of approval, then the participant begins to say the words and sentences required of them, at this moment we must mention that one of the most important things that we ask of the participants is to say the words in the required manner, meaning that they performed the acting and lived the role as if they were in an emergency or a natural state, and say the words or sentences with right feelings and tone.

3.2.6. Naming the Samples Files

Now, after these previous steps, we have one last step left in the method of collecting data which is naming files, and it is considered a very important step, in which we follow a unified method in naming files, and it contains several things.
• Collector: the name of the data collector, like Ali.
• Mic: the microphone device in: Arduino, mobile, or any other microphone
• Date: the date when the data was collected in the year-month-day format.
• Hour: the time when the data was first collected in the 24-hour format.

For example: Ali_Arduino_2023-02-03_16-30.wav
We also added some information to the files, Such as:
• Ambient: the type of the place when the data was collected, such as: noiseless, classroom, car, coffeeshop, street, or beach.
• Gender: the gender of the participant, such as: male or female.
• Nationality: the nationality of the participant, like Saudi.
• Languages: the participant languages separated by dashes, like: Arabic, Urdu-English, Bengali-Arabic-English.
After this step, we can say that we have ready-made high-quality data for analysis, and this leads us to the next section. At the end, the aggregated waveforms were renamed and split into training and test sets for data anonymization, as explained in the next section.

3.3. Data Analysis Method

We will talk here about the method we used to analyze the data we collected, and the first thing we do before starting the analysis process is to upload the data to the Edge Impulse platform. Now the data analysis steps begin.

3.3.1. Writing Label for Each Sample

In our case, we have two types: positive and negative. Positive is the data that we want the device to interact with and take a reaction, and the negative is the data that we want the device to consider as a natural event that does not require a reaction.

3.3.2. Segmenting the Samples

We had to set a fixed size segment for all samples, and the size was done in one second for each sample, which is the best size as indicated by we have two options to perform the segmentation process.

First Option: Do the segmenting process in the Edge Impulse platform, where we sample a long audio clip, for example: 10 seconds (in Data acquisition). There are options on the sample. Once we click on "Split sample", we will see something such as in Figure 24, we put the windows in the right places according to the audible sound. We can add other segments and move it to any place.

Second Option: we can also do the same thing with Audacity application.

Figure 24: Segmenting in Edge Impulse

3.3.3. Splitting the Data into Test and Training Samples

• What is Training Data?

Machine learning uses algorithms to learn from data in datasets. They find patterns, develop understanding, make decisions, and evaluate those decisions. In machine learning, datasets are split into two subsets. The first subset is known as the training data. It's a portion of our actual dataset that is fed into the machine learning model to discover and learn patterns. In this way, it trains our model.

• What is Testing Data?

Once your machine learning model is built with the training data, you need unseen data to test your model. This data is called testing data, and you can use it to evaluate the performance of the final product after adjusting the progress of your algorithms on the training.

• Why Knowing the Difference is Important?

The difference between training data versus test data is clear: one trains a model, the other confirms it works (or doesn't work) correctly with previously unseen data. However, confusion can pop up between the similarities and differences of both. And sometimes, at obviously AI, we'll see some people try and use their training data to make final model evaluation. This is why knowing the difference between the two is so important - you want to ensure you're fueling your models with the right data so you can get the best, most accurate insights. After all, those insights will feed directly into your decision-making. The training dataset is generally larger in size compared to the testing dataset. The general ratios of splitting train and test datasets are 80:20, 70:30, or 90:10.

Now that we've covered the differences between the two, we note that the training data can be split into two training and validation sets.

3.3.4. Impulse Design

In this step we add and modify some things so that we can take full advantage of the collected samples.

• Time series data: Here we set the window size that we want, in our case we want the window size to be a second.

• Adding a processing block: Extracting meaningful features from your data is crucial to building small and reliable machine learning models, and in Edge Impulse this is done through processing blocks. We ship a number of processing blocks for common sensor data (such as vibration and audio).

In our case, we are dealing with data from human voices. For this reason, we will use the MFCC features. MFCC is an audio feature extraction technique which extracts speaker specific parameters from the speech. It is the most popular and dominant method to extract cepstral features for speech by the use of perceptually based Mel spaced filter bank processing of the Fourier Transformed signal [17].

• Adding a learning block: After extracting meaningful features from the raw signal using signal processing, you can now train your model using a learning block. In our case, we use Keras Classifiers. Classification is a type of supervised machine learning algorithm used to predict a categorical label [18].

3.4. Details of Collected Data

We initially aspired to collect about 500 samples, but thank God we have collected 967 samples. The positive ones are 382, and the number of negative samples is 585. They were collected by the team members from 41 participants of different ages, genders, nationalities, and background voices. We tried to make as much as possible to work on diversifying the samples as much as possible because this contributes to raising the quality of the project's dataset. At the end of this chapter, we can now say that we have studied the data that we targeted, and after that we started the steps of complete technical preparation and legal preparation, such as preparing the participation form in the research and determining the type of words and sentences that we are required to collect, as well as the target groups of people that we are working on. Collecting data from them, then we uploaded the data to the Edge impulse platform, which enables us to analyze and classify the collected data. We expect the reader of this chapter to be able to fully follow the methodology we outlined in collecting data. We followed the academic

professional methods used in the most prestigious universities for the methodology of data collection, such as MIT, in order to obtain the best results, and to maintain the confidentiality and confidentiality of the collected data. We now expect that the road will be clear for those who want to work on developing the project, because we have developed a complete methodology and we have also built an initial database that facilitates work for those who want to take the project to another level.

4. Results

In this project, deep learning is the means to obtain results and based on this, there are major steps for deep learning. Starting with understanding the problem and defining the data, and we have already gone through in these phases. After collecting the data, it is time to train and test the model, which is one of the important steps in deep learning to get results. In this chapter, we will show you the results of our project step by step from machine learning to testing and obtaining results such as Accuracy and Loss.

4.1 Training and Testing the Model

After collecting data from 41 volunteers and 967 audio samples, the samples were divided into positive samples and negative samples, and each of the previous two sections was classified into two parts, a part for training and a part for testing. The main goal was to obtain the best possible accuracy. There was no specific target, and according to the results we extracted, we found that the best possible accuracy is 87.7%, with a loss of 0.34. All the results and changes that we made to the training process to reach this accuracy will be reviewed on these pages, with an explanation of the changes that occurred and the characteristics that were used in improvement and development.

• 1st Model:

In the first result, we use MFCC features with different layers and with 300 training cycles. Figure 25 shows the accuracy and loss of this Model. The matrix in Figure 26 explains the result of Negative Case & Positive Case when it was true or false in percentage. And we can also see the recall and the F1 scores of the samples, which are plotted in Figure 27. The result was obtained by the network of Figure 28.

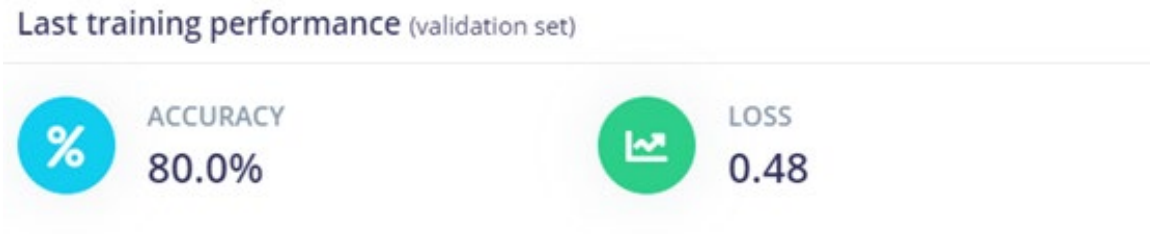

Figure 25: Accuracy & Loss of 1st Model

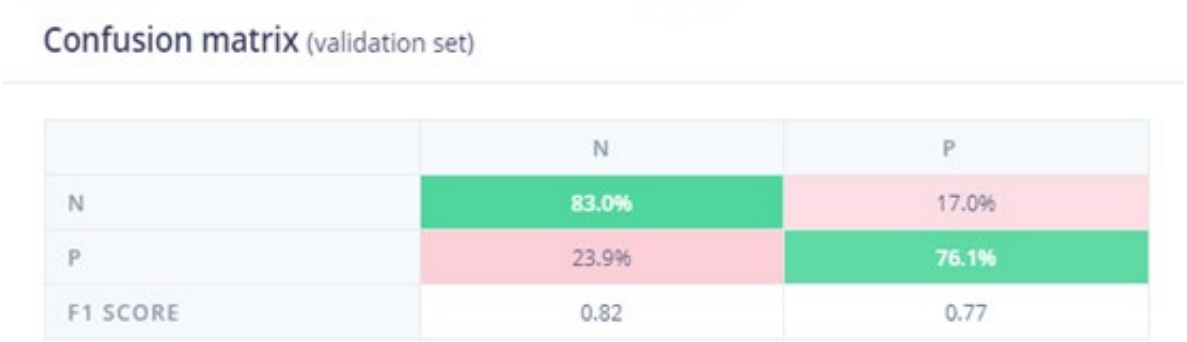

Figure 26: Confusion Matrix of 1st Model

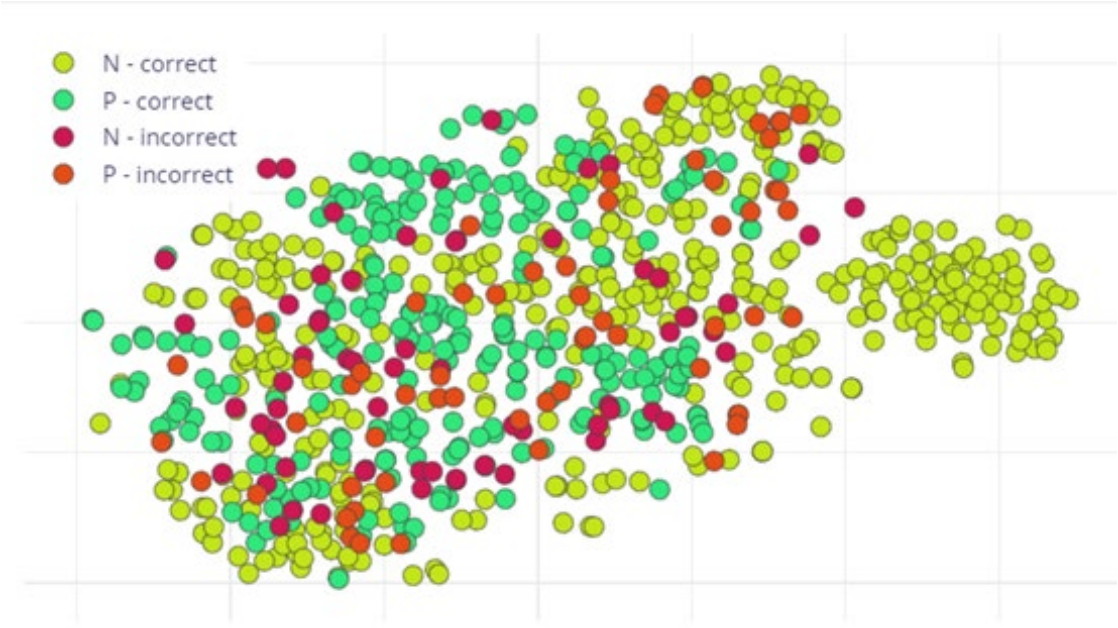

Figure 27: Data explorer of 1st Model

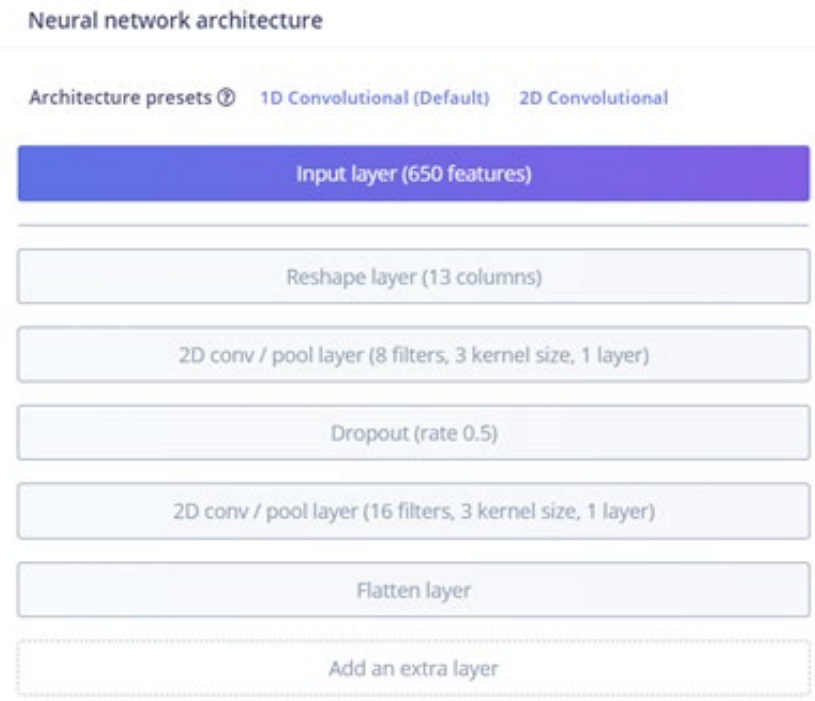

Figure 28: Neural Network Layers of 1st Model

• 2nd Model:

After the first results, we decided to improve the results by changing the MFCC features to MFE, changing the bugs for the neural network architecture, and trying to extract better results. And this is what happened, as the results appeared after this change with high accuracy, as shown in Figure 29 and Figure 30. The data and the model in this case are shown in Figure 31 and Figure 32.

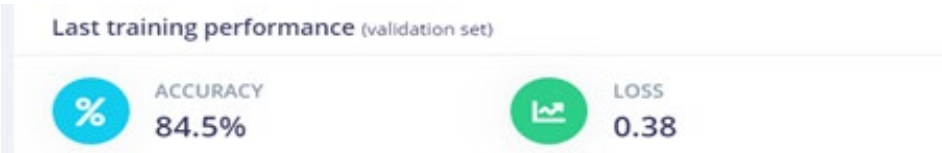

Figure 29: Accuracy & Loss of 2nd training

Confusion matrix (validation set)

	NEGATIVE CASE	POSITIVE CASE
NEGATIVE CASE	86.3%	13.7%
POSITIVE CASE	18.3%	81.7%
F1 SCORE	0.87	0.80

Figure 30: Confusion Matrix of 2nd Model

Figure 31: Data Explorer of 2nd Model

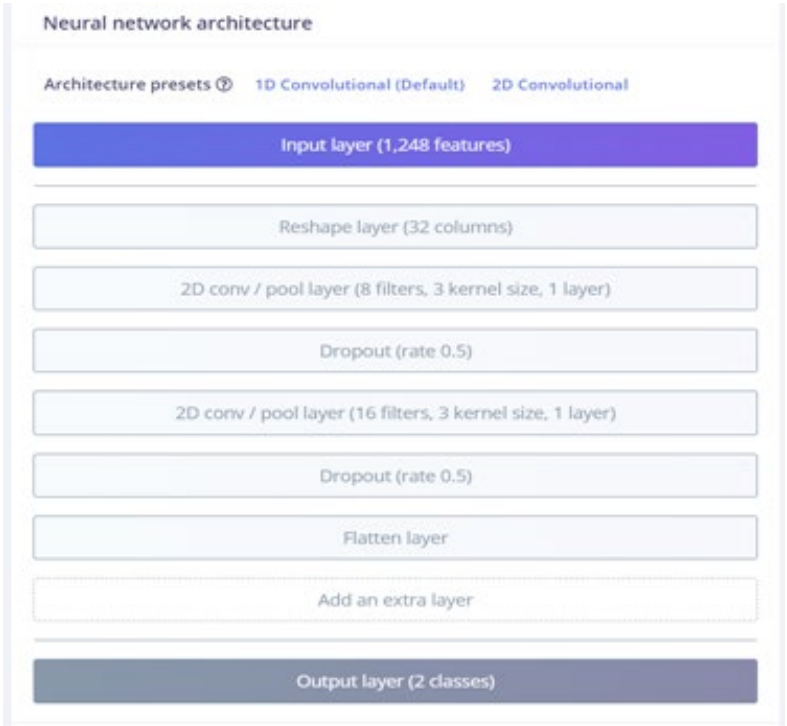

Figure 32: Neural Network of 2nd Model

In comparison between the first and second results, we find that the data did not change, but rather the change was in the properties and layers of the neural network, and this is a noticeable improvement, but we wanted more than that, and indeed we achieved a better result in the third Model thanks to the use of the EON tuner, which is available on Edge Impulse. Its function is to determine the best available options by searching to sort the models by their validation accuracy. This defines the characteristics of the classes for education in a very specific way. And after seeing several possible models, you can choose the characteristics that suit your requirements and the capabilities of your device, and in these pictures, I will show you some of the options that the fine-tuner provided to us for this project.

• Best EON Tuner Models:

Within the limitations of your target device, the EON tuner assists you in locating and choosing the optimum embedded machine learning model for your application. The tuner evaluates your input data, prospective signal processing pipelines, and neural network designs before providing you with a list of model architectures that might satisfy the latency and memory needs of the device you have selected. Figure 33 shows the interfaces of EON Tuner in edge Impulse with our project data and device, and you can compare between models.

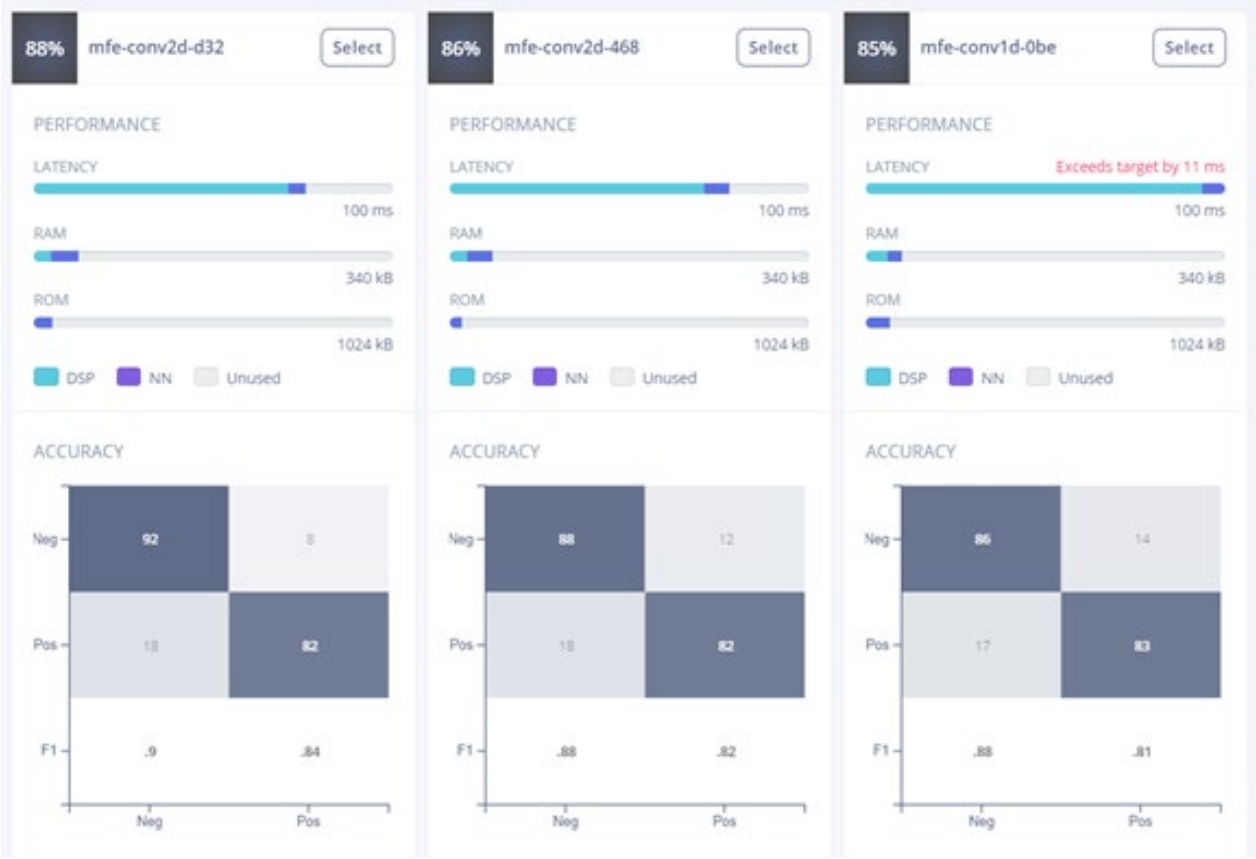

Figure 33: The best-embedded machine learning models for my Project

• 3rd Model

So, we chose the 3rd Model depending on the best-embedded machine learning model from the tuner, and Figure 34 shows the result of Accuracy and Loss. We can now see the improvement in accuracy and loss as well as the remarkable improvement in Figure 35 of the third model, whose scattering and model are in Figure 36 and Figure 37.

Figure 34: Accuracy & Loss for 3rd Model

Confusion matrix (validation set)

	NEGATIVE CASE	POSITIVE CASE
NEGATIVE CASE	86.3%	13.7%
POSITIVE CASE	18.3%	81.7%
F1 SCORE	0.87	0.80

Figure 35: Confusion Matrix of 3rd Model

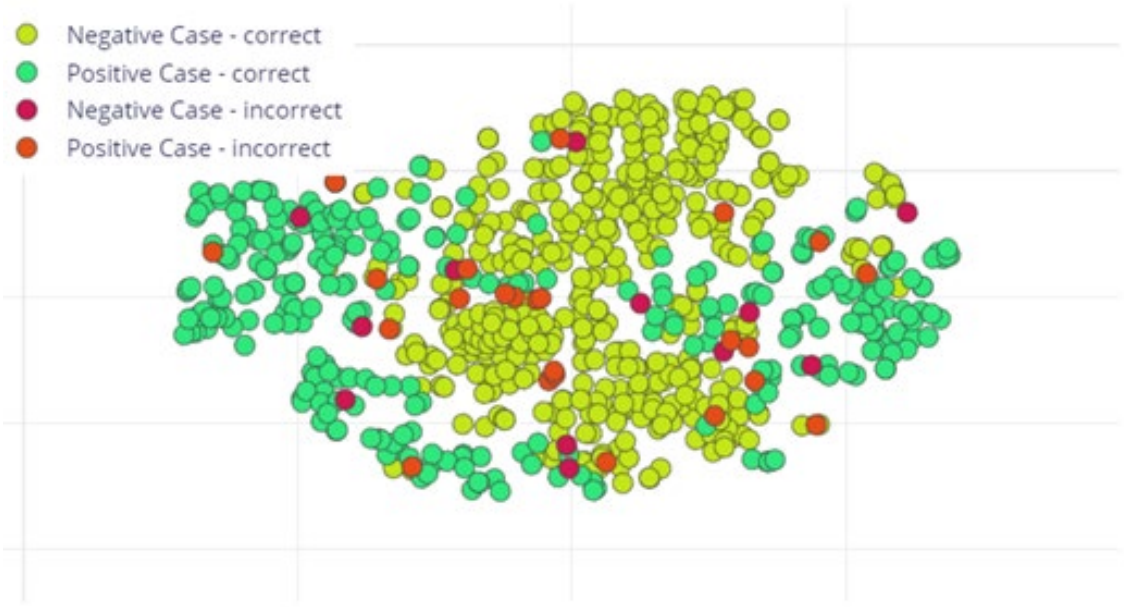

Figure 36: Data Explorer of 3rd Model

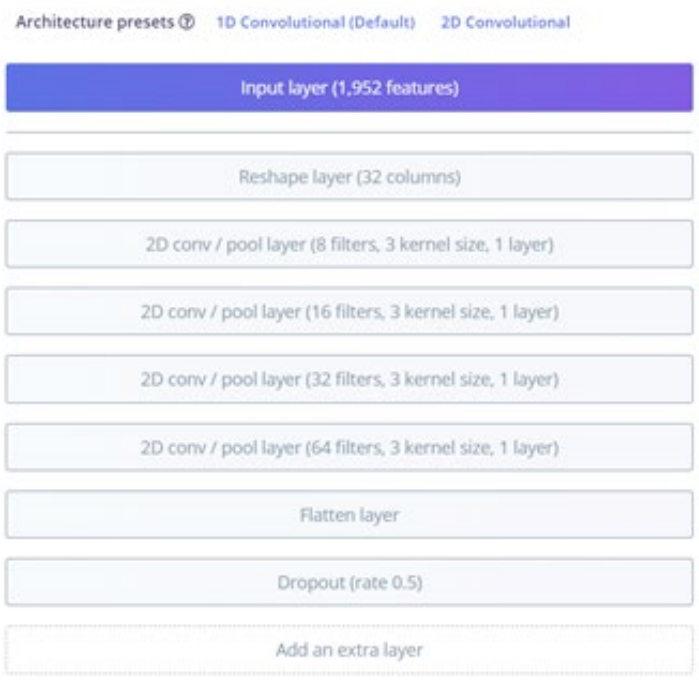

Figure 37: Neural Network Layers of 3rd Models

Finally, we implemented the final model in the physical microprocessor. Accuracy can also be further improved by increasing the number and variety of samples and determining the best fitting for neural network architecture. The graph in Figure 38 shows us that the best number of training circuits was 100 since we can see validation loss increases after 100 while the training loss decreases.

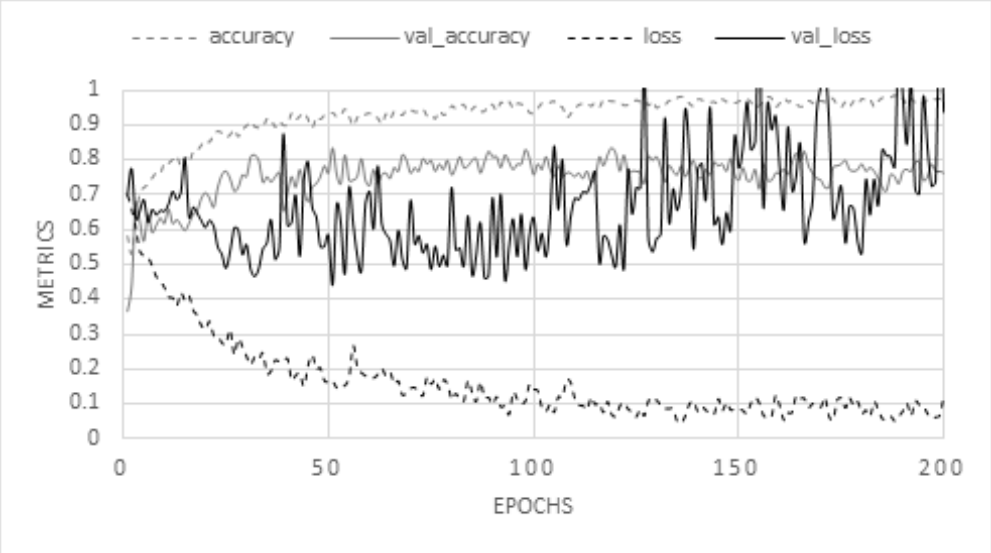

Figure 38: Early stopping to avoid overfitting the training progress

5. Conclusion

Emergency response strategies fall short of optimal processing and supply needs immediate help for people with special needs or a very important situation. In emergency situations, what people need most when they are faced with an emergency situation is a device that communicates their needs to their loved ones. Some emergencies make this difficult to make phone calls, therefore, vulnerable people can be helped by providing them with an intelligent system that can assess the dangers in the environment in a few seconds and alert the occupants of the place. Therefore, in this report, we have outlined a sound emergency alarm framework with deep learning of the Arabic language and the dialects of Saudi Arabia. Figure 39 shows validation accuracy and loss of three sets of hyper-parameters when the model was fine-tuned [19].

In this report, we touched on much important information for the pieces used, such as Arduino Nano 33 BLE Sense, Bluetooth, and sensors. As for the databases and samples used, data collection was utilized to ensure the availability of sufficient data and to ensure reliability in reporting the results and the conclusion, and we took samples through many volunteers, national and foreign, elderly and young, as well as men and women. A positive sample means that there is an emergency situation such as saying a fire, and negative samples are neutral samples that do not have an emergency situation, and they were taken in different dialects in order to increase the model's accuracy. Also, ideas that can be developed, such as sensors, were talked about by listing the nature of his work and its association with this project.

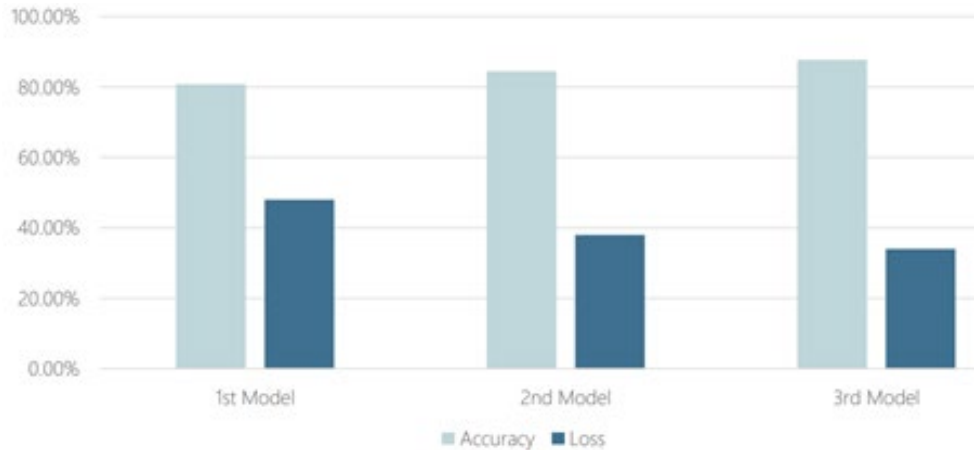

Figure 39: The Results During the Training Process to Reach this Accuracy

"An Emergency Alert System Based on Deep Learning in Arabic Language.

• My name is [**let the Child Know the Name**].
• We are asking you to participate in a research study because we are trying to create a device for "emergency alert in cases such as fire through speech," and we would like to record your voice in the audio device. We would like you to read the following sentences:

(Emergency Sentences)
• Fire
• Flames
• Help me, fire!
• Help me, Mom! Help me, Dad! Help me, Daddy!
• My father is in a fire! My dad is in a fire! My mother is in a fire! My mom is in a fire!

(Neutral sentences)
• I love my mother. I love my mom.
• There is no fire.
• There is no flames.
• The weather is very hot, like the heat of fire.

• If you agree to participate in this study, your voice will be stored in this device, and the device will recognize speech data in emergency situations.
• Your voice will be included in this device training data through your participation.
• By participating, you will help save others in emergency situations.
• Please discuss this matter with your parents before deciding whether or not to participate. We will also ask your parents for permission for you to participate in this study, but even if your parents say "yes," you can still choose not to participate.
• If you do not want to be in this study, there is no need to participate. Remember that your participation in this study is up to you, and no one will be upset if you do not want to participate or if you change your mind later and want to remove your voice.

English translation of the form

Agreement to participate in a research study titled:
"An Emergency Alert System Based on Deep Learning in Arabic Language."

1. My name is **[let the child know the name]**.
2. We are asking you to participate in a research study because we are trying to create a device for "emergency alert in cases such as fire through speech," and we would like to record your voice in the audio device. We would like you to read the following sentences:

(Emergency sentences)
- Fire
- Flames
- Help me, fire!
- Help me, Mom! Help me, Dad! Help me, Daddy!
- My father is in a fire! My dad is in a fire! My mother is in a fire! My mom is in a fire!

(Neutral sentences)
- I love my mother. I love my mom.
- There is no fire.
- There is no flames.
- The weather is very hot, like the heat of fire.

3. If you agree to participate in this study, your voice will be stored in this device, and the device will recognize speech data in emergency situations.
4. Your voice will be included in this device training data through your participation.
5. By participating, you will help save others in emergency situations.
6. Please discuss this matter with your parents before deciding whether or not to participate. We will also ask your parents for permission for you to participate in this study, but even if your parents say "yes," you can still choose not to participate.
7. If you do not want to be in this study, there is no need to participate. Remember that your participation in this study is up to you, and no one will be upset if you do not want to participate or if you change your mind later and want to remove your voice.

Figure 40: Children Volunteer Forms

Agreement to Participate in a Research Study Titled:
"An Emergency Alert System Based on Deep Learning in Arabic Language."

We are students from Umm Al-Qura University, from the College of Engineering in Al-Leith, Department of Electronics and Communications Engineering. We are working on a research project idea for an alarm device to detect emergency situations, such as cases of fire that operates through speech recognition. We are looking to collect voice samples from volunteers to contribute to the completion of this project. You must read the information below and ask questions about anything you do not understand before deciding whether or not to participate.

• Your participation in this research is voluntary.
• We want to collect audio samples to input into this device so that the device can recognize speech in emergency situations and save lives.
• You can withdraw your audio samples from the dataset at any time.
• You can refuse to participate.
• You can ask me questions about the project now or at any time.

Contact information:
Mobile: [phone number]
Email: [email]

You can also ask questions to the supervisor of this project, [supervisor's name].

Contact information:
Mobile: [phone number]
Email: [email]

Participation and Withdrawal:
Your participation is voluntary and you can withdraw at any time without penalty or consequences. For example, if you are a student from Umm Al-Qura University, it will not affect you if you refuse to participate in studies or other voluntary work.

Purpose of the Study:
We will create a device that helps others in emergency situations through sound.

English translation of the form

Agreement to participate in a research study titled:
"An Emergency Alert System Based on Deep Learning in Arabic Language."

We are students from Umm Al-Qura University, from the College of Engineering in Al-Leith, Department of Electronics and Communications Engineering. We are working on a research project idea for an alarm device to detect emergency situations, such as cases of fire, that operates through speech recognition. We are looking to collect voice samples from volunteers to contribute to the completion of this project. You must read the information below and ask questions about anything you do not understand before deciding whether or not to participate.

- Your participation in this research is voluntary.
- We want to collect audio samples to input into this device so that the device can recognize speech in emergency situations and save lives.
- You can withdraw your audio samples from the dataset at any time.
- You can refuse to participate.
- You can ask me questions about the project now or at any time.

Contact information:
Mobile: [phone number]
Email: [email]

You can also ask questions to the supervisor of this project, [supervisor's name].

Contact information:
Mobile: [phone number]
Email: [email]

- **Participation and withdrawal:**

Your participation is voluntary and you can withdraw at any time without penalty or consequences. For example, if you are a student from Umm Al-Qura University, it will not affect you if you refuse to participate in studies or other voluntary work.

- **Purpose of the study:**

We will create a device that helps others in emergency situations through sound.

Figure 41: Page 1 of the Adults Volunteer Forms

Procedures:
If you volunteer to participate in this study, we ask you to read the following sentence

- Sentences for an emergency event:
  - Fire.
  - I need help.

- Sentences for a natural event:
  - The game is on fire.
  - I saw a fire.
  - A fire occurred yesterday somewhere.
  - Has there been a fire accident in this place!

Potential Risks:
Your voice may be heard by other team members during programming or working on this device.

Expected Benefits to the Community:
By participating in this study, there may be a benefit for you and others.

Privacy and Confidentiality:
No identification about you will be disclosed to others, and the audio recording will be deleted upon your request.

Use of Data in the Future:
Your audio recording will not be used in other studies, and once the study is completed, all data used will be disposed of.

Consequences of Withdrawal:
There are no consequences for withdrawal from this study.

Volunteer Withdrawal from this Study:
If a volunteer withdraws from the study, it is considered their right and the team cannot publish any samples related to this person.

Purpose of Data Collection:
The purpose of data collection is to use it in an emergency disaster alert system based on deep learning, and the information you provide will only be available to organizations with access rights.

English translation of the form

- **Procedures:**

If you volunteer to participate in this study, we ask you to read the following sentences:

- Sentences for an emergency event:
  - Fire.
  - I need help.

- Sentences for a natural event:

1. The game is on fire.
2. I saw a fire.
3. A fire occurred yesterday somewhere.
4. Has there been a fire accident in this place!

- **Potential risks:**

Your voice may be heard by other team members during programming or working on this device.

- **Expected benefits to the community:**

By participating in this study, there may be a benefit for you and others.

- **Privacy and confidentiality:**

No identification about you will be disclosed to others, and the audio recording will be deleted upon your request.

- **Use of data in the future:**

Your audio recording will not be used in other studies, and once the study is completed, all data used will be disposed of.

- **Consequences of withdrawal:**

There are no consequences for withdrawal from this study.

- **Volunteer withdrawal from this study:**

If a volunteer withdraws from the study, it is considered their right and the team cannot publish any samples related to this person.

- **Purpose of data collection:**

The purpose of data collection is to use it in an emergency disaster alert system based on deep learning, and the information you provide will only be available to organizations with access rights.

Figure 42: Page 2 of the Adults Volunteer Forms

Duration of Retention of Samples:
This information may be retained for an indefinite period, and you have the right to withdraw your data from the study at any time. If you withdraw from the study, no new information will be collected about you by the research team.

By signing this form, I agree to participate in this study.

English translation of the form

- **Duration of retention of samples:**

This information may be retained for an indefinite period, and you have the right to withdraw your data from the study at any time. If you withdraw from the study, no new information will be collected about you by the research team.

By signing this form, I agree to participate in this study.

Volunteer name: _____
Signature: _____

Figure 43: Page 3 of the Adults Volunteer Forms

Ethics Approval and Consent to Participate
The study included audio data collection and was approved by the Electronics and Communications department chair of the College of Engineering at al-Lith.

Consent for Publication
All respondents consented for publishing their un-identifying participations by themselves or their legal guardians.

Availability of Data And Material
The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests
The authors did not have any financial and non-financial competing interests.

Funding
The authors did not receive any special funding.

Authors' Contributions
The document writing was contributed by the three two authors, the data collection was contributed by fourth and fifth authors, the research was supervised by sixth author, and it was reviewed by the last three.

Acknowledgement
We want to thank all the participants who cooperated with us to record their voice. Their efforts are appreciated and cared for by everyone and that the project would not have been completed without them.

References

1. D. P. Acharya. (2022). "6 Best Speech-to-Text API for Your Modern Applications," 19 12.
2. Arduino. (2023). "What is Arduino?".
3. "Nano 33 BLE Sense," Arduino.
4. https://www.sparkfun.com/products/17048
5. B. Dannegård. (2023). "Nano 33 BLE Sense Cheat Sheet," Arduino, 2 (20).
6. https://circuitdigest.com/microcontroller-projects/arduino-nano-33-ble-sense-board-%20%20%20%20review-and-getting-started-guide.
7. P. Inigo. (2008). 5 (5).
8. J. Bagur. (2023). "Connecting Nano 33 BLE Devices over Bluetooth®," 2 (20).
9. Chen, S. J., Hovde, D. C., Peterson, K. A., & Marshall, A. W. (2007). Fire detection using smoke and gas sensors. Fire Safety Journal, 42(8), 507-515.
10. Yu, L., Wang, N., & Meng, X. (2005, September). Real-time forest fire detection with wireless sensor networks. In Proceedings. 2005 International Conference on Wireless Communications, Networking and Mobile Computing, 2005. (Vol. 2, pp. 1214-1217). Ieee.
11. G. K. V. M. a. K. S. Venayagamoorthy. (1998). "Voice recognition using neural networks." Symposium on Communications and Signal Processing-COMSIG'98 (Cat. No. 98EX214). IEEE, 1998, South African.
12. Alu, D. A. S. C., Zoltan, E., & Stoica, I. C. (2017). Voice based emotion recognition with convolutional neural networks for companion robots. Science and Technology, 20(3), 222-240.
13. "GitHub," ShawnHymel, 26 3 2021.
14. Admin, "Research Methodology," 28 Mar 2022.

15. "About Edge Impulse," Edge Impulse.
16. K. Barked. (2022). "The Difference between Training Data vs. Test Data in Machine Learning," 11 February.
17. Sithara, A., Thomas, A., & Mathew, D. (2018). Study of MFCC and IHC feature extraction methods with probabilistic acoustic models for speaker biometric applications. Procedia computer science, 143, 267-276.
18. D. Singh. (2019). "Classification with Keras," 10 April.
19. https://circuitdigest.com/microcontroller-projects/arduino-nano-33-ble-sense-board-%20%20%20%20review-and-getting-started-guide.