Universiti Teknologi MARA

Home Safety Sound Detection System Using Convolutional Neural Networks

MUHAMMAD MUIZZUDDIN BIN ROSLAN

BACHELOR OF COMPUTER SCIENCE (Hons.)

Date

4th JULY 2024

SUPERVISOR APPROVAL

Home Safety Sound Detection System Using Convolutional Neural Networks

By

MUHAMMAD MUIZZUDDIN BIN ROSLAN

2022602128

This project was prepared under the supervision of the project supervisor, Zeti Darleena Binti Eri. It was submitted to the Collage of Computing, Informatic and Mathematics and was accepted in partial fulfilment of the requirement for the degree of Bachelor of Computer Science.

Approved by,

Zeti Darleena Eri

……………………….

Zeti Darleena Binti Eri

Project Supervisor

STUDENT DECLARATION

I certify that this thesis and the project to which it refers to the product of my own work and that any idea or quotation from the work of other people, published or otherwise are fully acknowledged in accordance with the standard referring practices of the disciplines.

…………………………

Muhammad Muizzuddin Bin Roslan

2022602128

Date Submit: 4th JULY 2024

ACKNOWLEDGEMENT

Alhamdulillah, praise and thanks to Allah because of His Almighty and His utmost blessing, I was able to finish this research within the time duration given. Firstly, my special thanks go to my supervisor, Zeti Darleena Binti Eri for guiding me throughout my journey to complete this research. Thank you for all your reprimands and guidance as it is very helpful for me in completing my report and also my system development.

Not to mention my heartfelt gratitude to both of my parents for always believing in me and being there for me from kindergarten to the end of my degree. I'm not sure where I'd be without both of you. Thank you for always providing me with motivation when I am at my lowest. I was incredibly blessed to have parents like you both.

Finaly, I'd like to express my gratitude to all my friends. Thank you always for being there for me throughout this degree process. Finally, I want to thank everyone who is assisting me in completing this research and system development.

Table of Contents

TABLE OF FIGURES

LIST OF TABLES

ABSTRACT

The goal of this project is to create a sound detection system for home safety that can detect sounds in domestic settings that may be dangerous or suspicious by using Convolutional Neural Network (CNN) techniques. The initiative, which takes cues from gadgets like Apple's iOS 14 accessibility tool and Amazon's Alexa, attempts to improve home security by precisely identifying alarms and shattering glass. The study highlights CNNs' superior performance in sound classification tasks, supporting their choice with empirical and literary evidence. Streamlit was used to construct an intuitive interface that accepts the entry of single, multiple, and real-time audio files. These files are then processed into spectrograms for analysis by CNN. With an overall classification accuracy of 96%, the MobileNetV2 model proved to be highly accurate in differentiating between different sound classes. Future research should focus on growing the dataset, maximizing noise reduction, boosting real-time processing, and improving user interface design for wider uptake. This study represents a significant leap in home security technology, demonstrating how deep learning, web-based interaction, and audio signal processing can be successfully integrated to produce a reliable sound detection system.

CHAPTER 1

INTRODUCTION

This chapter provides an overview of the research project, including the background study and the problem statement. This chapter also provides the project objectives, scope, and significance of Home Safety sound detection system to provide a better understanding of how to make this project a success.

1.1Background Study

During this year and before, many new applications or systems have been made by using sound detection. Sound detection is a process that detects any activities related to sound waves in the environment. This process can be used in many applications such as healthcare, security and environmental monitoring that improve people's lives in this modern day.

Statement made by Mark Purdy (2023), every year the U.S has spent billions in order to protect building and physical assets by creating smart home devices called Amazon's Echo. Presently, numerous artificial intelligence systems incorporate sensors and deep-learning algorithms capable of analyzing ambient sounds, and these technologies have been deployed in various settings to enhance safety measures.

Major brands have integrated sound detection technology for security purposes, exemplified by Amazon's Alexa. When users activate Alexa Guard before leaving their homes, Alexa-enabled devices can detect sounds such as glass breaking or smoke detectors going off, prompting users to receive notifications (Chieh-Chi Kao, 2020). Similarly, Apple introduced a new accessibility tool in iOS 14, enabling iPhones to detect 14 distinct sounds and alert their owners accordingly (Ashley Carman, 2020).

This project proposes the development of a system using Convolutional Neural Network (CNN) algorithms to identify potentially harmful or suspicious sounds within a domestic environment. The system is inspired by the sound detection technology integrated by prominent businesses such as Apple and Amazon. As an example, Chieh-Chi Kao (2020) notes that Amazon's Alexa Guard enables Alexa-enabled devices to detect sounds like glass breaking or smoke detectors going off and inform users accordingly. Similar to this, iPhones can recognize 14 different sounds and notify their owners thanks to Apple's accessibility feature in iOS 14 (Ashley Carman, 2020). In keeping with these illustrations, our solution seeks to improve home security through the identification of important noises and prompt notifications.

The rationale behind selecting Convolutional Neural Network (CNN) algorithms for this project stems from their proven superior performance and efficiency compared to previous methodologies in the field of environmental sound classification. Studies, such as that conducted by Abdoli Sajjad et al. (2019), have demonstrated the superiority of CNNs over alternative methods, including a "shallow" dictionary learning model with augmentation and a deep, high-capacity model without augmentation.

1.2 Problem Statement

The perception of home as a sanctuary of safety is common among people, offering comfort and a sense of distance from danger. However, it is important to recognize that homes can also be venues for potentially hazardous activities.

One such concern is the occurrence of burglary, which is a prevalent issue across various countries. For instance, statistics from the Bureau of Justice indicate that property crime rates in the United States stood at 32.5% and 33% in the years 2019 and 2020, respectively (Morgan et al., 2021).

Moreover, homes can harbor unforeseen dangers, particularly for vulnerable individuals such as elderly people living alone. Research indicates that in the United States, nearly 30% of the 60 million older citizens reside alone, with half of them being 85 years or older and about three-quarters being women (Daniel B. K., 2023).

In light of these risks, enhancing home security measures becomes imperative to ensure the safety of occupants. For instance, the installation of a comprehensive safety system can significantly mitigate potential threats, thereby warranting consideration for implementation in this proposed project.

When considering the improvement of safety and security within a household setting, the adoption of Convolutional Neural Networks (CNNs) is justifiable due to their efficacy in detecting sounds and identifying anomalies. CNNs have proven to be highly successful in diverse tasks, such as classifying environmental sounds and detecting unusual sounds, rendering them well-suited for integration into home security systems.

In the article "Anomalous Sound Detection Based on Convolutional Neural Network and Mixed Features" authored by Jie Zhao (2020), the focus lies on highlighting the proficiency of CNNs in precisely discerning abnormal sounds. This capability holds significant importance in identifying unforeseen and potentially hazardous activities within the confines of a household.

1.3. Objectives

The objectives of the project are as shown below:

- To investigate the elements of potential threats within the application of CNN algorithms for Home Safety.
- To implement a home safety sound detection system using Convolutional Neural Network (CNN) algorithms.
- To evaluate the home safety system, use CNN algorithm accuracy and efficiency.

1.4. Project Scope

For this project scope, it will mainly be focusing on sound detection.

i. The target user
   The target audience for this system will be homeowners or individuals with authority over the premises.

ii. Data.

Data that will be collected and use for the training for this project is in a form of audio file.

iii. Algorithm.

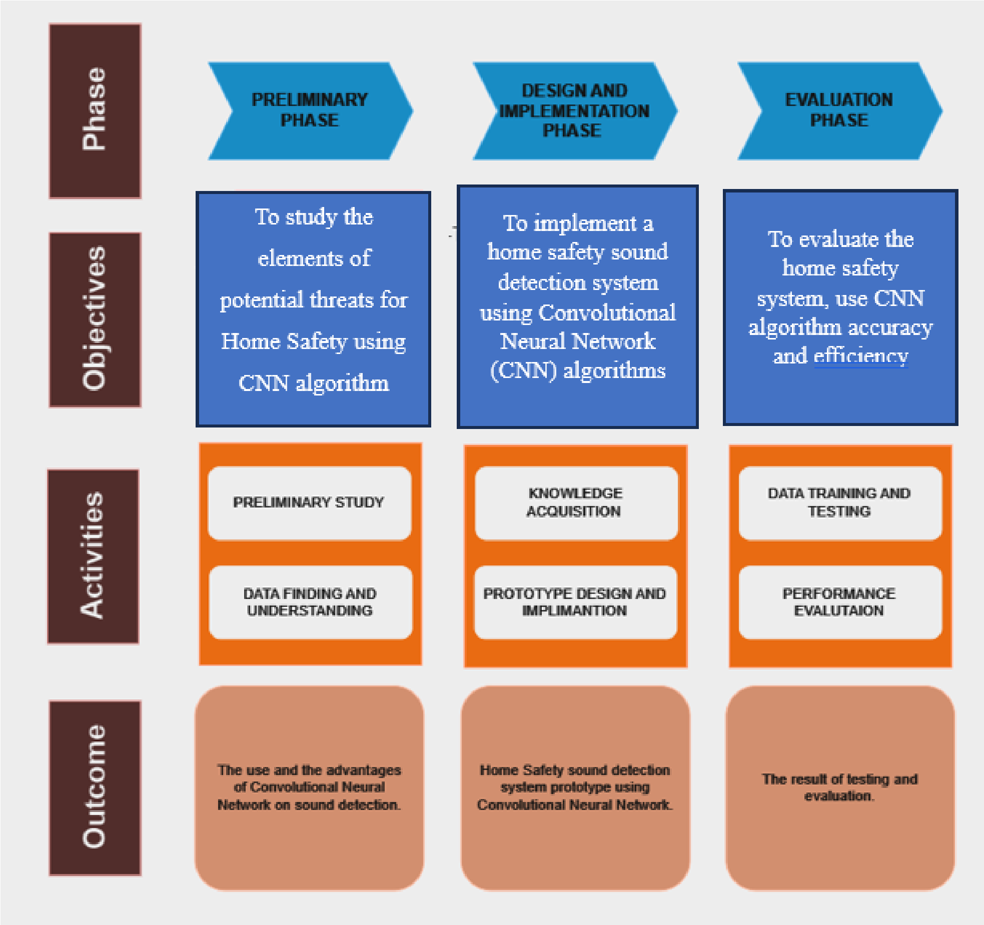
The algorithm that will be used for this project is Convolutional Neural Network (CNN). CNN is an algorithm that uses image processing.

1.5 Project Significance

The home safety sound detection system offers invaluable assistance to citizens by promptly detecting and alerting them to potential safety and security hazards such as glass breaking, verbal aggression, or gunshots. This proactive approach enables individuals facing imminent danger to take preemptive action, whether by summoning authorities to their location or preparing for self-defense.

Furthermore, through continual training and experience accumulation, the system can enhance its effectiveness and accuracy in identifying sounds occurring in the vicinity, ultimately enabling its seamless integration into real-time applications.

1.6 Overview of Research Framework



Phase	PRELIMINARY PHASE	DESIGN AND IMPLEMENTATION PHASE	EVALUATION PHASE
Objectives	To study the elements of potential threats for Home Safety using CNN algorithm	To implement a home safety sound detection system using Convolutional Neural Network (CNN) algorithms	To evaluate the home safety system, use CNN algorithm accuracy and efficiency
Activities	PRELIMINARY STUDY DATA FINDING AND UNDERSTANDING	KNOWLEDGE ACQUISITION PROTOTYPE DESIGN AND IMPLIMANTION	DATA TRAINING AND TESTING PERFORMANCE EVALUTAION
Outcome	The use and the advantages of Convolutional Neural Network on sound detection.	Home Safety sound detection system prototype using Convolutional Neural Network.	The result of testing and evaluation.

Figure 1. 1 Overview of Research Framework.

The research framework comprises three distinct phases. The preliminary phase involves studying Convolutional Neural Network (CNN) and acquiring relevant data, culminating in an understanding of the utilization and benefits of CNN algorithms for sound detection.

Following this, the design and implementation phase will focus on activities such as acquiring knowledge, designing prototypes, and implementing the system. The outcome of this phase will be the development of a Home Safety sound detection system employing Convolutional Neural Network (CNN) technology.

Finally, the evaluation phase will entail data training, testing, and performance evaluation. The outcomes of this phase will include the results of testing and the overall evaluation of the system. Figure 1.1 illustrates the overview of the research framework described above.

1.7 Conclusion

In conclusion, this study conducted is made to create a home safety sound detection using Convolutional Neural Network. This project's main objective was to study the work of Convolutional Neural Network in sound detection. The project target user will be the homeowner or the people with the authorities on the building, so that people can have sigh of relief for their safety.

CHAPTER 2

LITERATURE REVIEW

A literature review is a critical and thorough analysis of published academic works, books, research articles, and other pertinent materials that are either directly or indirectly connected to a certain topic or question of study. It provides a synthesis of the present state of knowledge in a given topic and acts as a fundamental component of academic and research initiatives. A literature review serves several purposes, including highlighting the approaches and conclusions of earlier studies, constructing the background and theoretical framework for a new study, and identifying gaps and limits in the body of current research. By means of this procedure, scholars get a refined comprehension of the past and present discussions pertaining to their selected subject, empowering them to situate their research within the wider scholarly context. In addition to providing guidance for the researcher's strategy and methods, a well-written literature review advances knowledge by developing, analysing, and extending the concepts found in the body of current literature.

2.1Background

People could experience a range of difficulties about their safety in their homes. These difficulties may arise from a variety of sources, such as one's own actions, the outside world, and the physical surroundings.

To tackle these obstacles, a proactive and comprehensive strategy is needed. By being aware of potential risks, taking preventative action, doing routine safety inspections, and being ready for emergencies, people can improve their safety at home. A safer living environment can also be achieved by encouraging a culture of safety within the home and getting expert help when necessary.

2.1.1 Overview of Home Safety

Ensuring home safety requires a comprehensive assessment of potential hazards within and around the household. This evaluation encompasses various aspects, including injury risks, combustion safety, pest issues, and cleanliness. Individuals can conduct a home safety audit independently by identifying common household dangers and regularly monitoring them (Tholen, 2021).

Home safety strategies aim to safeguard individuals of all ages and abilities within their living environments. These strategies involve identifying risks to the safety of children, infants, individuals with physical or cognitive impairments, and aging adults who opt to remain in their homes (Davis, 2023). Implementing safety techniques includes hazard removal, installation of safety devices, and educating both patients and caregivers on safety measures (Davis, 2023).

2.1.2 The Importance of Home Safety Sound Detection

The potential for improving people's overall security and well-being in a residential setting makes home safety sound detection with Convolutional Neural Networks (CNNs) an important application. Deep learning models such as CNNs are ideally suited for sound detection applications in the context of home safety, as they have demonstrated efficacy in a range of audio-related tasks.

In terms of home safety, using Convolutional Neural Networks for sound detection improves security, automates monitoring, and issues early alerts for any threats. By combining CNNs with smart home technology, homeowners can create a more responsive and intelligent living space that is ultimately safer for them.

2.2 Sound Detection

2.2.1 Introduction

The process of sensing and transmitting sound waves in the surrounding environment is known as sound detection. Applications for this technology are numerous and include consumer electronics, healthcare, industrial monitoring, security systems, and consumer electronics. As Halil Ozgen Dindar et.al (2021) said in their article the used of sound detection is growing day by day because its possibility to be used in wide area of places. The goal of sound detection is to basically to analyse and interpret audio signals, making it possible to identify occurrences, trends, or anomalies.

As the technology keep on growing, based on Future Market Insights (2023) in their report Sound Sensor Market, the revenue in 2022 was US$ 1,412.2 Million and expected to reach US$ 2,426.5 Million by 2023, as it is estimated to grow at a Compound Annual Growth Rate (CAGR) of 5.1% for 2023-2033. This could conclude that in coming years there will be more research or technology that will be relay based on sound detection into various application.

More complex sound detection systems have also been developed as a result of technological advancements. For example, a study on the application of machine learning to identify and characterize the sound produced by fish demonstrates the use of deep learning algorithms for fish sound detection (Barroso et.al, 2023). In the upcoming years, it is anticipated that these developments in sound detection technologies will keep propelling the expansion of the sound sensor and audio recognition industries.

2.2.2 Technologies utilised

The sound detection technologies of 2018 and 2023 have progressed considerably. A few of the essential tools and methods are as follows:

1. **Machine Learning and Deep learning (Method):** Both techniques have been used in various areas, including detection, classification, and identification of biological acoustic signals (Barroso et.al, 2023). For supervised learning tasks in sound identification, machine learning techniques such random forests (RF), K-nearest

neighbors (KNN), support vector machines (SVM), logistic regression, naive Bayes, and linear regression have been employed (Barroso et.al, 2023). For more difficult tasks, deep learning techniques such as recurrent neural networks (RNN) and convolutional neural networks (CNN) have also been used (Barroso et.al, 2023).

2. **Data Processing and Preprocessing (Method):** The creation of effective preprocessing and data processing methods is essential for managing big datasets and enhancing the functionality of sound detection systems (Barroso et.al, 2023). Sound detection models have been optimized by applying various techniques such matching filter, spectrogram correlation, energy thresholds, and data augmentation to determine detection thresholds (Barroso et.al, 2023).
3. **Voice Recognition Technology (Tools):** Voice recognition technology is a computing approach that allows specific software and systems to identify and verify each speaker's unique voice (DigitalJournal, 2023). It's a technology that lets machines understand and recognize human speech and convert it into text or commands. Numerous industries, including banking, telecommunications, healthcare, the government, and consumer applications, have used this technology (Zhang X.,2023).

2.2.3 Challenges and Limitation

There are several obstacles in the way of developing automatic systems for sound event recognition, some of which are connected to the types of sounds that need to be identified. The variety and intricacy of real-world sound settings are one of the main obstacles to sound detection. The variety of sounds found in the environment can make it challenging to create sound detection systems that can precisely recognize and categorize a broad range of sounds (Mesaros, 2021). Furthermore, sound recognition algorithms have considerable hurdles due to the dynamic nature of sound environments and the presence of background noise. These algorithms must be robust enough to discern between relevant noise and target sounds (Mesaros, 2021).

The requirement for big and varied labelled datasets to train machine learning models is another barrier to sound detection. Large-scale sound dataset acquisition and annotation can be labour-intensive and time-intensive processes, especially for uncommon or specialized sound events. The effectiveness and generalizability of sound detection systems may be restricted in the absence of high-quality labelled data, which is essential for training precise sound detection models (Mesaros, 2021).

2.3 Algorithms

2.3.1 Convolutional Neural Network (CNN)

2.3.1.1 Introduction of Convolutional Neural Network

One kind of deep learning architecture that has become very popular is the Convolutional Neural Network (CNN), especially for image recognition and computer vision applications. Deep learning, a larger family of machine learning techniques based on learning data representations, includes CNNs as a subset (IBM, 2023).

Due to their ability to learn the spatial hierarchies of features automatically and adaptively from the input data, CNNs are utilized for computer vision and image recognition tasks. This is accomplished by applying fully connected, pooling, and convolutional layers, which allow the network to recognize patterns in the input data (IBM, 2023). The most important component of a CNN is the convolutional layer, which uses a series of filters to the input data to produce feature maps that identify the existence of particular patterns or features (IBM, 2023).

2.3.1.2 Overview of Convolutional Neural Network

Deep learning architectures known as convolutional neural networks (CNNs) have attracted a lot of interest, especially in the areas of image recognition and computer vision applications. They belong to the deeper family of machine learning techniques called "deep learning," which is centred on learning data representations (IBM, 2023).

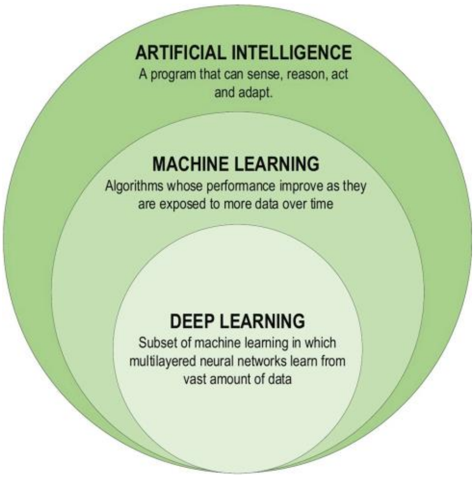

ARTIFICIAL INTELLIGENCE
A program that can sense, reason, act and adapt.
MACHINE LEARNING
Algorithms whose performance improve as they are exposed to more data over time
DEEP LEARNING
Subset of machine learning in which multilayered neural networks learn from vast amount of data

Figure 2. 1 Hierarchy of Artificial Intelligence (Alzubaidi et.al, 2021).

In 2012, Krizhevsky A, Sutskever I, and Hinton GE introduced AlexNet, marking a significant milestone in the advancement of deep convolutional neural networks (CNNs), as noted by Alzubaidi et al. (2021). Their research paved the way for remarkable progress in image recognition and classification, notably demonstrated in the ImageNet Large Scale Visual Recognition Challenge. By employing various parameter optimization techniques and deeper network architectures, AlexNet substantially improved the learning capabilities of CNNs.

Furthermore, Hinton's suggestion to randomly pass over many transformational units during training to ensure robust feature learning was used by Krizhevsky et al. to counteract overfitting, a negative associated with higher depth, according to Alzubaidi et al. (2021) in their article. In addition, the vanishing gradient issue was addressed, and the convergence rate was increased by using the rectified linear unit (ReLU) as a non-saturating activation function.

In order to reduce overfitting and enhance generalization, overlapping subsampling and local response normalization were also used. To improve the network's performance over past architectures, the scientists also made changes such using large-size filters (5 × 5 and 11 × 11) in the initial layers (Alzubaidi et.al, 2021).

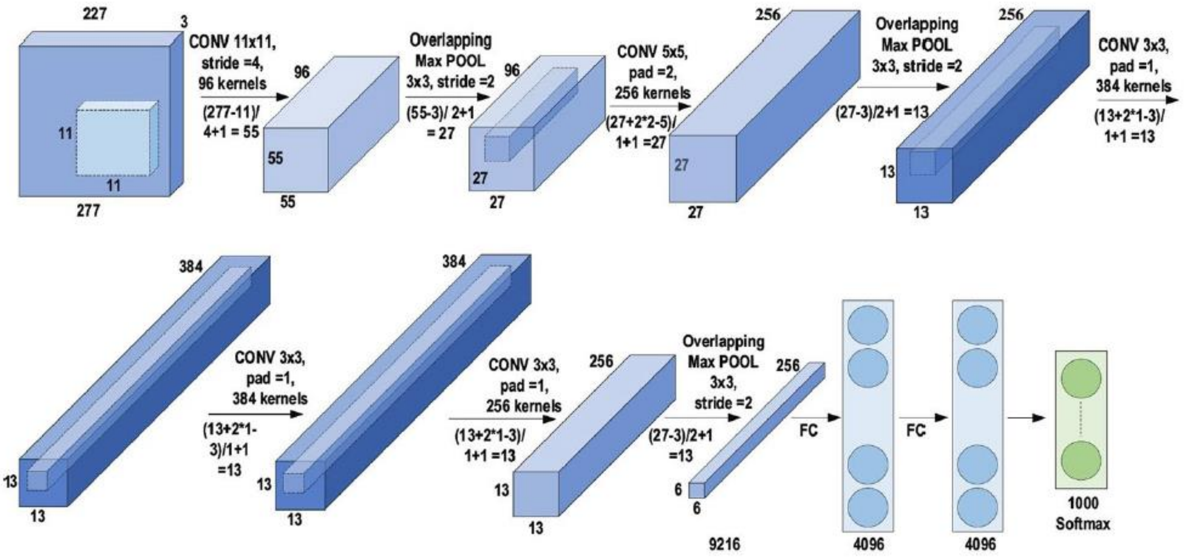


Figure 2. 2 Architecture of AlexNet (Alzubaidi et.al, 2021).

CNNs have applied many applications, according to Alzubaidi et.al (2021) in their article such as:

1. **Image classification:** Image classification refers to the process of categorizing and labeling groups of pixels or vectors within an image according to specific criteria. This categorization is typically based on one or more spectral or textural characteristics (Boesch G., 2021). While convolutional neural networks (CNNs) are currently considered the most advanced approach for image classification, various other machine learning algorithms and deep learning techniques have also been utilized in this field over time (Boesch G., 2021).
2. **Object detection:** Object detection in computer vision involves the identification and localization of objects within images or videos. This capability finds applications across a wide range of fields including surveillance, autonomous driving, and medical imaging (Qureshi R., 2023). Recent advancements in deep learning, particularly the adoption of convolutional neural networks (CNNs) and algorithms like You Only Look Once (YOLO), have significantly enhanced the performance of object detection systems (Qureshi R., 2023).
3. **Image segmentation:** Image segmentation refers to the process of assigning labels to each segment or region of an image, effectively partitioning it into smaller parts. This task is crucial in numerous applications such as object recognition, remote sensing, and medical imaging (Drew Banks, 2023). The development of more precise and efficient

image segmentation models, capable of real-time segmentation, has been made possible by advancements in deep learning techniques (Drew Banks, 2023).

2.3.1.3 Advantages and Disadvantages of Convolutional Neural Network

According to Alzubaidi et.al (2021), CNNs have their own advantages and disadvantages. Table 2.1 below show the advantages and disadvantages of Convolutional Neural Network (CNN).

Table 2. 1 Advantages and Disadvantages of CNNs.

Advantages	Disadvantages
Sparse Connectivity: CNNs exhibit sparse connectivity, implying that neurons within one layer are only partially connected to a few neurons in the subsequent layer. This characteristic reduces the number of connections and weights required, thereby enhancing CNNs' processing efficiency for large datasets (Alzubaidi et al., 2021).	**High-quality labelled data**: Acquiring high-quality labelled data poses a challenge for effectively training Convolutional Neural Networks (CNNs), as they heavily rely on extensive datasets. Consequently, this reliance may lead to less precise outcomes (Alzubaidi et al., 2021).

Table 2.1 Advantages and disadvantages of CNNs. (Continued)

Advantages	Disadvantages
Weight sharing: CNNs minimize the number of parameters needing to be learned by employing the same set of weights across the entire input image. Through weight sharing, CNNs achieve translation-invariant properties, enabling the network to recognize objects regardless of their location within the image (Alzubaidi et al., 2021).	**Large Memory Requirements**: CNNs demand significant memory resources, particularly when handling large datasets and complex architectures. Storing network weights and activations necessitates ample memory, which can escalate computational expenses and sluggishness during both training and inference, notably on low-power devices (Alzubaidi et al., 2021).
Improved Performance: Studies have demonstrated superior performance of CNNs over traditional neural networks in tasks	**Limited Interpretability**: The interpretability of CNNs is often limited, as they are sometimes perceived as opaque

related to image and video processing, particularly in areas such as object detection, image recognition, and video analysis (Alzubaidi et al., 2021).	models. Understanding the rationale behind the network's predictions can prove challenging, posing significant issues in scenarios where transparency and accountability are paramount (Alzubaidi et al., 2021).
Hierarchical Feature Learning: CNNs adopt a hierarchical approach to feature learning, progressively advancing from basic features like edges to more complex features such as shapes and objects. This hierarchical feature learning enables CNNs to grasp the underlying structure of input data and generate accurate predictions (Alzubaidi et al., 2021).	**Limited Generalization**: While CNNs excel at processing images and videos, their ability to generalize to novel or untested data may be restricted. This limitation is particularly evident in cases of bias or inadequate training data, leading to overfitting and inferior performance on fresh datasets (Alzubaidi et al., 2021).

2.3.2 Long Short-Term Memory Networks (LSTMs)

2.3.2.1 Introduction of Long Short-Term Memory Network

LSTMs, or Long Short-Term Memory networks, represent a type of RNN, or recurrent neural network, capable of preserving long-term dependencies in sequential data, as highlighted by Mayank Banoula (2023). Sequential data such as text, audio, and time series can be effectively processed and interpreted by LSTMs. These networks address the vanishing gradient problem commonly encountered in conventional RNNs by regulating information flow through the use of gates and a memory cell, thus enabling them to selectively retain or discard information as required (Mayank Banoula, 2023). Widely utilized across various domains, LSTMs find applications in tasks such as time series forecasting, speech recognition, and natural language processing (Mayank Banoula, 2023).

2.3.2.2 Overview of Long Short-Term Memory Network

The Long Short-Term Memory (LSTM) represents a specific category of recurrent neural networks (RNNs) designed to handle extended dependencies within sequential input. LSTMs demonstrate proficiency in processing various types of sequential data such as text, audio, and time series, as highlighted by Mayank Banoula (2023).

Comprising a sequence of LSTM cells, the architecture of an LSTM network is structured to manage long-term dependencies effectively (Mayank Banoula, 2023). Within each cell, a collection of input, output, and forget gates serves to regulate the flow of information entering and exiting the cell and these gates play a crucial role in the LSTM's ability to retain relevant information from previous time steps while selectively discarding unnecessary data, as elaborated by Mayank Banoula (2023).

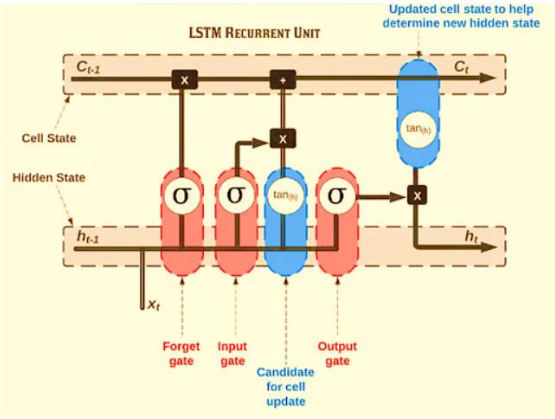


Figure 2. 3 Structure of LSTM, (Mayank Banoula, 2023).

Time series forecasting, speech recognition, and natural language processing are just a few of the fields and businesses in which long-term memory banks (LSTMs) are used. They have showed amazing success in various tasks, typically exceeding regular RNNs and other machine learning methods (Mayank Banoula, 2023).

2.3.2.3 Advantages and Disadvantages of Long Short-Term Memory Network (LSTMs)

According to Kousias K. (2020), Sugandhi A. (2023) and the web site geeksofgeeks (2020), LSTMs have both advantages and disadvantages of their own. Table 2.2 below show the advantages and disadvantages of LSTMs based on their research.

Table 2. 2 Advantages and disadvantages of LSTMs.

Advantages	Disadvantages
Time series data's capacity to identify long-term dependencies (Kousias K., 2020).	Significant resource requirements and time commitment for preparation (geeksforgeeks, 2020).
Excellent for mobile broadband network forecasting tasks, for example (Kousias K., 2020).	System capabilities are frequently not met by high memory bandwidth requirements (geeksforgeeks, 2020).
Realizes cutting-edge bandwidth forecasting capabilities, especially in 5G settings (Kousias K., 2020).	Although being built to overcome it, struggles with long-term dependencies in the input data (Sugandhi A., 2023).
Performance can be greatly impacted by hyperparameter changes (Kousias K., 2020).	Inability to understand time relationships longer than a few steps (Sugandhi A., 2023).

2.4 Implementation of Convolutional Neural Network in Various Problem

Convolutional Neural Networks (CNNs) are the preferred solution for a variety of problems, especially those involving image recognition and computer vision tasks. This is because CNNs are very effective at tasks involving visual perception and pattern recognition because they can automatically learn hierarchical features, recognize spatial patterns, and generalize well to new and unseen data.

Prachi Juyal and Amit Kundaliya (2023) conducted a study titled "Multilabel Image Classification using the CNN and DC-CNN Model on Pascal VOC 2012 Dataset" with the aim of addressing the challenging task of multilabel picture annotation in computer vision. Their research endeavors to enhance the performance of automated photo annotation by introducing a novel dual-channel convolutional neural network (DC-CNN) model specifically designed for multilabel image categorization. In this study, convolutional neural networks (CNNs) and DC-CNNs were employed as the primary methodologies. CNNs were utilized to discern patterns in pixel images, while the newly proposed DC-CNN model was introduced to augment the effectiveness of multilabel image categorization.

The dataset utilized by the authors for their investigation is The Pascal VOC 2012 dataset, renowned for its comprehensive collection of 20 distinct categories. To facilitate their research, the dataset was partitioned into three distinct subsets: a private testing set, comprising an unspecified number of data points, a validation set consisting of 1,449 data points, and a training set comprising 1,464 data points. The authors chose to leverage the Pascal VOC 2012 Dataset to both train and evaluate their proposed DC-CNN model for multilabel image classification.

In terms of accuracy, the findings of the study indicate that the DC-CNN approach achieved an impressive average maximum accuracy (AP) of over 95% for the training data class, utilizing solely the Pascal VOC 2012 dataset. Moreover, the proposed CNN technique demonstrated exceptional performance, achieving a perfect accuracy rate of 100% for the training data class. These results underscore the efficacy of the DC-CNN model in enhancing multilabel image classification accuracy, showcasing promising advancements in the field of computer vision research.

There also other research study than were use as reference for this project. Table 2.2 below show other implementation of Convolutional Neural Network (CNN) algorithms in various type of problem.

Table 2. 3 Implementation of CNNs in various types of problems.

No.	Title	Objective	Algorithms	Problems	Advantages	Disadvantage	Accuracy	Citation
1	Detection and Classification of Lung Abnormalities by Use of Convolutional Neural Network (CNN) and Regions with CNN Features (R-CNN)	To create and assess computer-aided diagnosis (CAD) algorithms for lung issues, utilizing Convolutional Neural Networks (CNNs) and Region-based CNN features (R-CNN). The aim is to enhance CAD system precision in identifying and categorizing lung nodules and diffuse lung diseases in radiological images, aiding radiologists in their diagnostic processes.	Convolutional Neural Network and image -based CADe algorithm that use R-CNN	Accurately and efficiently diagnosing lung issues like nodules and diffuse diseases poses challenges, requiring specialized training and expertise from radiologists.	The use of convolutional neural networks (CNN) and regions with CNN features (R-CNN) allows for image-based computer and aided diagnosis (CAD) and detection algorithms that do not require an image-feature extractor, which can be difficult to design for complicated image patterns of lung diseases.	The dataset used in the research was relatively small, which may limit the generalizability of the results to larger and more diverse populations.	For the image-based CADx algorithm, the authors reported an accuracy of 91.7% in classifying lung nodules as benign or malignant and for the image-based CADe algorithm, the authors reported a sensitivity of 96.3% and a specificity of 98.4% in detecting lung abnormalities	Kido S. et.al (2018)

Table 2.3 Implementation of CNNs in various problem. (Continued)

No.	Title	Objective	Algorithms	Problems	Advantages	Disadvantage	Accuracy	Citation
2	Comparative Investigations on Tomato Leaf Disease Detection and Classification Using CNN, R-CNN, Fast R-CNN, and Faster R-CNN	to help farmers identify plant diseases. to evaluate the performance of the proposed technique using the Plant Village dataset and compare it with other deep learning algorithms.	Convolutional neural networks (CNNs), region-based CNNs (R-CNNs), fast R-CNNs, and faster R-CNNs	the early detection and classification of tomato leaf diseases that led to losses in the agricultural sector.	a user-friendly application that can assist farmers in identifying tomato leaf diseases, the afflicted areas, the accuracy of the disease diagnosis, and the treatments available.	The author does not specifically talk about any disadvantages.	an accuracy of more than 98% using a Faster R-CNN and VGG 16, which is a promising result for disease identification.	G. Priyadharshini, & Dolly, J. (2023)

Table 2.3 Implementation of CNNs in various problem. (Continued)

No.	Title	Objective	Algorithms	Problems	Advantages	Disadvantage	Accuracy	Citation
3	The Sparsity and Activation Analysis of Compressed CNN Networks in a HW CNN Accelerator Mod	to analyse the sparsity and activation of compressed CNN networks in a HW CNN accelerator model to increase sparsity through CNN compression and evaluate the performance of the compressed CNN networks by applying them to a CNN HW accelerator model.	Convolutional Neural Network (CNN)	the high computing power and bandwidth requirements of CNN networks.	it focuses on improving the efficiency and processing time of CNN networks, which can be a significant challenge in various applications that utilize CNN networks.	the study focuses primarily on the effects of CNN compression on sparsity and activation analysis, without considering other factors such as energy consumption or memory usage	The author does not specifically talk about any detail about the accuracy, but they manage to improve the efficiency and processing time of CNN networks.	Lee M. Y. et.al (2019)

Table 2.3 Implementation of CNNs in various problem. (Continued)

No.	Title	Objective	Algorithms	Problems	Advantages	Disadvantage	Accuracy	Citation
4	A Study on Object Detection Method from Manga Images using CNN	to explore the effectiveness of different object detection methods for panel layout, speech balloon, character face, and text in manga images. to compare the detection results of detectors trained using the same dataset for Fast R-CNN, faster R-CNN, and SSD, and to examine the change of detection rate for comic images by region proposals.	Fast R-CNN, Faster R-CNN, and SSD.	the detection of multiple object classes in manga images, which is complicated by the dense arrangement of manga objects.	the use of a large and diverse dataset of manga images, the comparison of three different object detection algorithms, and the examination of the effectiveness of object proposals for manga object detection.	it focuses solely on manga images, which may limit its generalizability to other types of images.	According to the findings, Faster R-CNN had the highest mAP of 0.91, followed by SSD with 0.86 and Fast R-CNN with 0.87.	Yanagisawa H. et al. (2018)

Table 2.3 Implementation of CNNs in various problem. (Continued)

No.	Title	Objective	Algorithms	Problems	Advantages	Disadvantage	Accuracy	Citation
5	Multilabel Image Classification using the CNN and DC-CNNModel on Pascal VOC 2012 Dataset	to propose a dual-channel convolutional neural network (DC-CNN) model to improve the efficiency and accuracy of multilabel image classification. to demonstrate the effectiveness of the DC-CNN model on the Pascal VOC 2012 dataset and to provide insights into the potential of deep learning models for multilabel image classification.	dual-channel convolutional neural network (DC-CNN) model.	Automated image labelling, particularly in multilabel image classification, presents a challenge due to variations in object appearance, pose, and illumination. This complexity, coupled with a limited amount of available training data, makes picture categorization difficult.	The article suggests that the training period for the proposed DC-CNN model is quick, which could have implications for real-time or near-real-time multilabel image classification applications and using the popular Pascal VOC 2012 Dataset, the article ensures evaluating and comparing the performance of the proposed model against established benchmarks in computer vision.	While widely used, dependence on the Pascal VOC 2012 Dataset may introduce limitations like class imbalance, restricted diversity, or biases, potentially affecting the generalizability of the DC-CNN model across various multilabel image classification tasks and real-world scenarios.	The proposed DC-CNN approach achieved an average maximum accuracy (AP) of above 95% for train data class on just the Pascal VOC 2012 data sets. Additionally, for the train data class, the suggested CNN approach achieves 100% accuracy.	Prachi Juyal, & Amit Kundaliya. (2023)

Table 2.3 Implementation of CNNs in various problem. (Continued)

No.	Title	Objective	Algorithms	Problems	Advantages	Disadvantage	Accuracy	Citation
6	Discovering the Optimal Setup for Speech Emotion Recognition Model Incorporating Different CNN Architectures	to discover the optimal setup for a speech emotion recognition model incorporating different CNN architectures. to improve the performance of different architectures in natural speech databases and to determine the best CNN architecture that can provide much better accuracy than the generated result in this study	Convolutional Neural Network (CNN)	the problem of accurately recognizing emotions from speech signals.	it explores the effectiveness of different deep learning approaches, such as 1D CNN, 2D CNN, and CNN LSTM, on a natural database.	The author does not specifically talk about any disadvantages.	the researchers concluded that there is still more work needed to discover in order to improve the given architecture for persuasively recognizing the emotions.	Joshua P. et al. (2022)

Table 2.3 Implementation of CNNs in various problem. (Continued)

No.	Title	Objective	Algorithms	Problems	Advantages	Disadvantage	Accuracy	Citation
7	TimeDistributed-CNN-LSTM: A Hybrid ApproachCombining CNN and LSTM to Classify BrainTumor on 3D MRI Scans PerformingAblation Study	to develop a deep learning model for accurate classification of brain tumors on 3D MRI scans. to achieve higher accuracy in tumor classification. to mimic the analysis pattern of radiologists and medical experts to develop an accurate, reliable, and effective performance that can outperform other existing approaches.	This model combines Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM)	the problem of accurate classification of brain tumors on 3D MRI scans	The model is designed to mimic the analysis pattern of radiologists and medical experts, which can lead to an accurate, reliable, and effective performance that can outperform other existing approaches	the proposed model is tested on a limited number of datasets, and the results may not be generalizable to other datasets.	the proposed TD-CNN-LSTM model outperforms all the studies with a test accuracy of 98.90%.	Montaha S. et al. (2022)

Table 2.3 Implementation of CNNs in various problem. (Continued)

No.	Title	Objective	Algorithms	Problems	Advantages	Disadvantage	Accuracy	Citation
8	A comprehensive survey of the R-CNN family for object detection.	to offer a comprehensive survey of deep learning R-CNN techniques and to present a comparison between those techniques in terms of test time per image, speed, and mean average precision (mAP). to provide a starting point for researchers to understand current research and identify open challenges for future research in the field of object detection using deep learning R-CNN techniques.	Region Based Convolutional Neural Networks	problems related to object detection in various fields such as computer vision, robotics, and autonomous systems.	it provides a starting point for researchers to understand current research and identify open challenges for future research in the field of object detection using deep learning R-CNN techniques.	it mainly focuses on the R-CNN family of algorithms and does not cover other object detection techniques, such as YOLO or SSD. Therefore, the research may not provide a complete picture of the state-of-the-art in object detection using deep learning techniques.	The author does not specifically tell the detail of the accuracy, but they manage to improve the accuracy and detection speed compared to R-CNN and Fast R-CNN	O. Hmidani, & Ismaili, M. (2022)

2.5 Similar Work on Home Safety Sound Detection

Similar initiatives have been investigated in research on home safety sound detection, looking into things like smart house technologies, surveillance systems, and sensor-based solutions to improve security and reduce dangers in home surroundings. These related studies highlight a concerted effort by the academic and technological community to improve the accuracy and resilience of Home Safety sound detection.

1."Prevention of safety accidents through artificial intelligence monitoring of infants in the home environment" is the title of a 2019 study by Lee Y. et al. that aims to create an intelligent infant monitoring system that can recognize risky scenarios and notify caregivers in the house. The suggested approach makes use of CNN to identify the baby's face and body, which are crucial markers of their state of sleep. The suggested system detects the infant's body and face in real-time using a deep learning algorithm called YOLOv3-tiny, which is a variation of the YOLO (You Only Look Once) method.

A sizable dataset of photos showing babies in various sleeping positions and environments is used to train the algorithm. After identifying the baby's body and face, the algorithm evaluates the photos to determine how well the child is sleeping. The system will determine that the baby is sleeping safely, for instance, if their face is pointed toward the camera and their body is in a safe position. However, the system will determine that the baby is in a risky scenario if their body is in an unsafe position or if their face is covered.

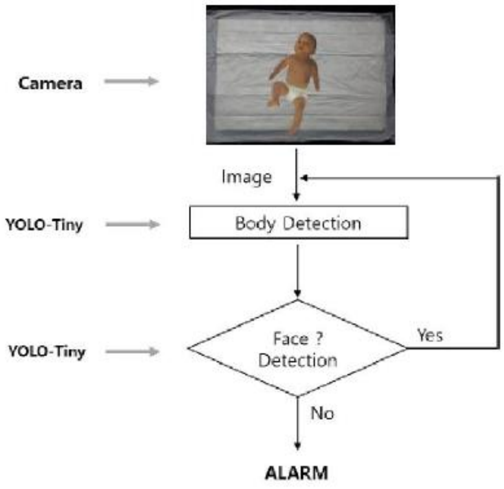
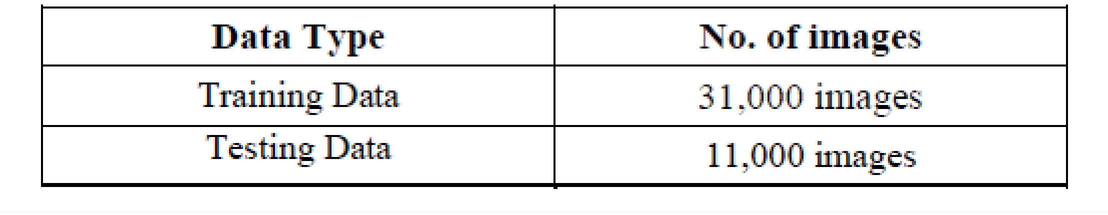


Figure 2. 4 The System Flowchart Lee Y. et al. (2019)

The system will promptly notify caretakers via a notification on their smartphone or other device if it senses that the infant is in a risky condition, such as suffocating or falling. This enables babysitters to act quickly to stop mishaps and guarantee the baby's safety.

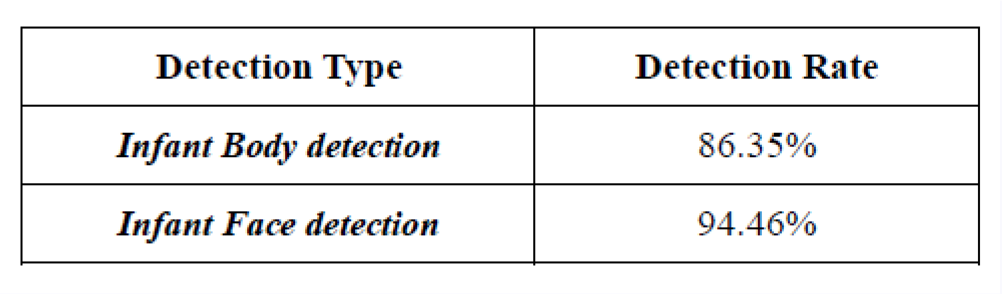
To assess the system's performance, the researchers used a virtual dataset in addition to a dataset consisting of 31,000 photos from ETRI. The achieved accuracy for face detection was 94.46%, while the accuracy for body detection was 86.35%. These outcomes show how well the suggested approach can identify and evaluate the sleeping conditions of infants.

Table 2. 4 Training and Testing Data by Lee Y. et al. (2019)



Data Type	**No. of images**
Training Data	31,000 images
Testing Data	11,000 images

Table 2. 5 The Result of infant Detection by Lee Y. et al. (2019)



Detection Type	**Detection Rate**
Infant Body detection	86.35%
Infant Face detection	94.46%

Ultimately, the goal of the research is to create an intelligent monitoring system that may avoid accidents and instantly notify caregivers of any changes in the home environment. This would help solve the pressing problem of safety mishaps involving newborns.

Table 2. 6 Similar works on Home Safety Sound Detection.

No.	Title	Objective	Algorithms	Problems	Advantages	Disadvantage	Accuracy	Citation
1	Prevention of safety accidents through artificial intelligence monitoring of infants in the home environment	to propose an intelligent infant monitoring system that can prevent safety accidents involving infants and children in the home environment. to evaluate the performance of the proposed system through experiments and achieve high accuracy in face and body detection.	You Only Look Once (YOLO)	the need for an intelligent infant monitoring system to prevent safety accidents involving infants and children in the home environment.	The paper evaluates the performance of the proposed system through experiments and achieves high accuracy in face and body detection.	Although the paper mentions the use of two datasets for training and testing the proposed system, the total number of images used is relatively small (31,000 pieces of image data for learning and 11,000 pieces of image data for detection performance test). This may limit the generalizability of the proposed system to other datasets.	The research achieves high accuracy in face detection rate is 94.46%, and the body detection rate is 86.35%.	Lee Y. et al. (2019)

Table 2.6 Similar works on Home Safety Sound Detection. (Continued)

No.	Title	Objective	Algorithms	Problems	Advantages	Disadvantage	Accuracy	Citation
2	Behavior-Rule Specification-based IDS for Safety-Related Embedded Devices in Smart Home	to propose a behavior-rule specification-based Intrusion Detection System (IDS) to mitigate security attacks on safety-related embedded devices in the smart home environment. to derive behavior-rules for smart home limited to safety-related sensors to provide security suitable for resource-constrained IoT-assisted smart homes.	Support Vector Machine (SVM) and Decision Tree (DT)	the security concerns that arise with the increasing use of IoT devices in building a safe home environment. the limitations of existing Intrusion Detection Systems (IDS) in supporting security in resource-constrained IoT-assisted smart homes and proposes a behavior-rule specification-based IDS to provide security suitable for the said environment.	The proposed IDS is specifically designed for the resource-constrained nature of IoT-assisted smart homes, making it suitable for deployment in such environments. The IDS is based on behavior-rule specification, which allows it to detect known threats and their variants, as well as protecting against unknown threats or zero-day attacks.	The proposed IDS is limited to safety-related sensors in the smart home environment, which means that it may not be suitable for detecting other types of security threats. - The performance of the proposed IDS was only validated in MATLAB tool, which may not accurately reflect its performance in real-world deployment scenarios.	The AUROC of the proposed IDS is close to 100%, indicating that the false positive probability and false negative probability are close to 0%.	Keon Yun et al. (2021)

2.6 Implication of Literature Review

A thorough analysis of the literature on Convolutional Neural Networks (CNNs)-based Home Safety Sound Detection Systems can have a lot of beneficial effects for the field's advancement. First off, a detailed analysis of the literature offers valuable insights into the most recent advancements in sound detection technologies, approaches, and issues pertaining to home safety applications. With the use of this knowledge, researchers and practitioners can expand on previously established bases, possibly advancing the development and improvement of CNN-based algorithms for increased accuracy and performance.

The literature review may also reveal creative methods and fresh concepts that could motivate the creation of sound detection models with higher efficacy. Through the identification of effective implementations and optimal methodologies, the review advances our collective comprehension of the practical applications of CNNs and provides invaluable direction for forthcoming research initiatives. A thorough evaluation of the literature also promotes a greater comprehension of the constraints and difficulties that current home safety sound detection systems must overcome. This helps researchers fill in knowledge gaps and identify new directions for investigation.

A survey of the literature in this area can have a number of beneficial effects, such as enhancing knowledge, spurring creative thinking, and providing a roadmap for improving CNN-based algorithms for sound detection in home safety.

2.7 Conclusion

Convolutional Neural Networks (CNNs) are used in Home Safety Sound Detection Systems. The literature analysis on these works concludes with a persuasive picture of the state of advancements and applications in the field of smart home technology. The review of previous studies highlights how important CNNs are for improving the precision and effectiveness of sound detection systems, especially when it comes to home security.

The researched literature highlights the adaptability and effectiveness of CNNs and offers a thorough grasp of the potential and constraints related to their implementation in diverse contexts. The knowledge gathered from this review of the literature helps to identify best practices and possible areas for Home Safety Sound Detection System design and implementation to be improved. Furthermore, the paper demonstrates how CNNs may be tailored to a wide range of problem domains, demonstrating their efficacy in tasks like image identification, natural language processing, and medical diagnosis.

This highlights the wide-ranging influence and suitability of CNNs, not only for sound identification systems but also for tackling intricate problems in several domains. CNN integration into Home Safety Sound Detection Systems is positioned to significantly advance the development of intelligent and secure living environments as researchers and practitioners expand upon the discoveries reported in the literature.

CHAPTER 3

RESEARCH METHODOLOGY

The research methodology for the Home Safety Sound Detection System utilizing the Convolutional Neural Network (CNN) algorithm is a crucial aspect of this study. This chapter will provide an in-depth exploration of the steps and methods employed to identify, select, organize, and assess information pertinent to the development of the system. The discussion will break down the research process into stages, elucidating each phase's significance in the creation of an effective sound detection system. The use of CNNs for sound categorization and the methodical development procedure that guarantees the effectiveness and functioning of the Home Safety Sound Detection System will be highlighted. This chapter will work as a thorough guide, illuminating the approaches used to answer the research questions and objectives, facilitating transparency into the choices taken during the development process, and providing insights into the general architecture and operation of the system. This research seeks to advance the fields of sound detection technologies and home safety by providing an in-depth analysis of the research process and methods.

3.1 Overview of Research Methodology Framework

This project technique uses the Waterfall Model, an organized, step-by-step process for developing systems. This three-phase approach guarantees a methodical and sequential development from the initial ideation to the last assessment of the Home Safety Sound Detection System.

The following components make up the framework:

1. Preliminary Phase:
   Problem definition, literature review, information acquisition, and objective formulation are the main priorities of the preliminary phase. A comprehensive literature study is

carried out to investigate previous efforts on sound detection systems and Convolutional Neural Networks (CNNs), and the issue statement is clearly defined. Information is gathered from a variety of sources, and clear goals are developed to direct the next steps.

2. Design and Implementation Phase:

   The phase that separates preparation from execution is known as design and implementation. System design includes developing a logistic regression pseudocode, user interface design, system flowchart, and comprehensive architecture. After that, the prototype is put into practice using the design guidelines. Clarity and precision are promoted by the Waterfall Model's linear progression, which guarantees that each component is fully developed before going on to the implementation phase.

3. Evaluation Phase:

   In order to evaluate the effectiveness of the Home Safety Sound Detection System, the evaluation phase is essential. Thorough testing, validation, and performance assessment are part of this step. Evaluation metrics are used to quantify how well the system performs in accomplishing its goals. These measures include detection accuracy, expert-based assessments, and overall system efficiency. Because of its organized methodology, the Waterfall Model enables a thorough examination prior to system finalization.

Because the Waterfall Model is sequential, there is less chance of ambiguities and doubts because each phase is clearly progressed through. This study technique was selected to offer a transparent and methodical framework for the development and assessment of the Home Safety Sound Detection System because of its clearly defined stages.

3.1.1 Detailed Content of Research Methodology Framework

There are several phases in the methodology framework for developing the Home Safety Sound Detection System, starting with the preparatory phase. The framework's first phase entails creating a clear problem description, conducting a comprehensive literature study, and gathering pertinent information. The observed gaps and insights are then used to methodically design the objectives.

The next step is the data collection, which entails carefully obtaining pertinent data that is appropriate for analysis.

As the framework moves into the design phase, it entails the development of a strong system architecture, an in-depth system flowchart, and an improved user interface design. Additionally, the system's technological base is strengthened by the incorporation of logistic regression pseudocode.

The software and hardware recommendations for an efficient and successful Home Safety Sound Detection System are combined in the prototype implementation phase that follows.

In the evaluation phase, a thorough performance review and assessment criteria such as detection accuracy and expert-based evaluations are introduced.

The documentation phase also includes a detailed project timeline and milestone tracking, a description of the prototype, and information on the testing and evaluation procedures. From conception to implementation and evaluation, this structured methodology framework guarantees a methodical and comprehensive development approach for the Home Safety Sound Detection System.

Table 3. 1 Detailed Content of Research Framework.

Research Methodology/Phase	Objective	Description	Task	Activities	Deliverable
Preliminary Phase	to study the elements of potential threats for home safety using CNN (Convolutional Neural Network) algorithm.	Perform an in-depth review of literature focusing on the application of CNN algorithms in home safety system.	Literature Review	Reviewing scholarly journals, articles, and books pertaining to collaborative filtering and recommender systems. Evaluating online materials, research papers, and case studies. Recognizing the methodologies and techniques employed in analogous projects.	Conduct an investigation into the historical context of collaborative filtering and recommender systems. Identify current challenges and propose solutions. Establish project goals associated with CNN and home safety systems. Examine the importance of CNN in enhancing home safety systems.

Table 3.1 Detailed Content of Research Framework (Continued)

Data Collection	to study the elements of potential threats for home safety using CNN algorithm.	Collect pertinent data to develop a thorough comprehension of the CNN algorithm within a home safety system.	Data collection	Do online searching on kaggles website to collect appropriate data that will be used to train and test the system.	Relevant datasets for CNN algorithm research.
Design Phase	to implement a home safety sound detection system using CNNs algorithm	Develop a suitable design plan for implementing CNNs algorithms in the home safety sound detection.	System design.	Designing structure and functionalities of the Home Safety sound detection system. Incorporating CNNs algorithm into the system. Defining user interface.	Detailed system structure for home safety sound detection system with CNNs algorithm. A complete design of the system with a appropriate user interface.

Table 3.1 Detailed Content of Research Framework (Continued)

Development phase	to implement a home safety sound detection system using CNNs algorithm	Implement the designed home safety sound detection system with CNNs	System development.	Implementing coding and programming according to the design specifications. Incorporating CNNs into the system. Executing iterative testing and debugging processes.	Functional Home Safety sound detection system with CNNs.
Testing phase	to evaluate the home safety system, use CNN algorithm accuracy and efficiency	Evaluate the Home safety system and the accuracy and the efficiency through testing.	System Testing	Conducting performance testing for the system accuracy and the efficiency in detecting sound.	The response from the system.
Documentation	N/A	Prepare documentation for the project	Documentation	Creating technical documentation for system architecture, algorithms used, and implementation details	A finished report of the proposed project

3.1.2 Significance of Choosing Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a highly significant choice for the creation of a Home Safety Sound Detection System because of their exceptional capacity to extract complex patterns and features from audio data. This is especially important when attempting to identify possible safety hazards in a home setting. CNNs are adept in processing and learning hierarchical representations from sound spectrograms, allowing for the extraction of complicated features indicative of various safety-related audio occurrences, such as glass breaking or alarms.

They are highly suitable for sound classification tasks due to their ability to detect spatial patterns in both the frequency and time domains, which guarantees precise and effective detection of important auditory information. CNNs can help the Home Safety Sound Detection System identify and classify a wider range of sound occurrences, which will strengthen and bolster the system's ability to warn occupants of possible safety hazards.

Additionally, CNNs help the system adapt to various acoustic settings and locations, guaranteeing its usefulness and efficacy in actual home safety situations. In general, the importance of selecting Convolutional Neural Networks stems from their ability to handle intricate audio information, making it possible to create an intelligent and adaptable Home Safety Sound Detection System that enhances people's safety and wellbeing in their homes.

3.2 Preliminary Phase

3.2.1 Problem Statement Formulation

Although many people believe that their houses are safe havens, they can also serve as locations for dangerous acts like burglaries. For example, according to Morgan et al. (2021) the Bureau of Justice, property crime rates in the US were 32.5% in 2019 and 33% in 2020. Furthermore, half of the 60 million elderly Americans who live alone are 85 years of age or older, accounting for nearly 30% of the population (Daniel B. K., 2023).

Existing sound detection systems can be dangerous to residents since they frequently can't tell the difference between normal household noises and possible security threats. Therefore, the development of a Home Safety Sound Detection system that employs Convolutional Neural

Networks (CNNs) to precisely identify and categorize auditory cues associated with safety hazards is imperative.

CNNs are perfect for home security systems since they have shown success in tasks like identifying anomalies and categorizing sounds from the environment. By resolving these problems, a dependable, real-time sound detection system can be created to improve home security and provide homeowners peace of mind.

3.2.2 Literature Review

A thorough examination of previous research in the fields of sound detection systems and CNN applications across a range of issue domains is included in the literature review on home safety sound detection and CNN implementation. The development of sound detection systems for home safety has been the subject of numerous research, all of which have stressed the significance of promptly identifying key auditory events like glass shattering, alarms, or strange sounds. In an effort to improve these systems' precision and effectiveness, research in this field has looked into cutting-edge technology such as CNNs for reliable pattern detection and classification.

Simultaneously, an abundance of literature has been devoted to the flexible uses of CNNs in solving various issues. In several different fields, including image identification, natural language processing, medical diagnostics, and others, CNNs have shown impressive results. The literature demonstrates CNNs' versatility across several problem domains and their capacity to automatically acquire hierarchical features from complex data, which makes them appropriate for challenges involving spatial patterns.

The synthesis of these literature streams informs the proposed Home Safety Sound Detection System's framework, demonstrating that the addition of CNNs can considerably boost the accuracy and reliability of sound detection in a home situation. Through the utilization of established techniques and comparable works, this literature evaluation provides a basis for the creation of a novel and efficient system for home safety through sophisticated auditory monitoring.

3.2.3 Knowledge Acquisition

Convolutional neural networks (CNNs) are used in home safety sound detection systems. The process of acquiring knowledge for these systems include obtaining, organizing, and combining pertinent data in order to improve the system's capacity to recognize and categorize safety-related sounds. In this setting, information is gathered from a variety of sources, such as academic papers, domain experts, and datasets with annotated audio samples. The first step in the process is to identify important sources of information, such as sound detection experts, previously published studies on acoustic event categorization, and databases containing sound samples that have been annotated in relation to home safety.

The collecting step entails methodically compiling data regarding various auditory occurrences that are important for home security, like glass breaking, alarms, and strange noises. Understanding the complexities of auditory patterns relevant to home safety is made possible by interviews with specialists in sound detection and a thorough analysis of the literature. Additionally, the CNN model is trained and validated using datasets that comprise sound samples related to safety issues.

When newly collected information is organized and prepared to meet the specifications of the CNN model, representation is put into action. In this step, sound data is transformed into spectrograms or other formats that can be used to train neural networks. Expert knowledge about the characteristics that set apart noises linked to safety is converted into attributes that the CNN can learn from.

Validation guarantees the dependability and correctness of the learned material. This entails validating with specialists, comparing the obtained data with established conclusions, and evaluating the datasets' quality. The goal is to ensure that the information incorporated into the system aligns with domain knowledge and previous research.

The last stage is integration, in which the learned material is smoothly integrated into the Home Safety Sound Detection System. The CNN model is trained using the structured data, expert insights, and annotated datasets, which help the machine learn and identify audio patterns linked to safety.

By using CNNs to provide precise and effective sound categorization in a home setting, knowledge acquisition in this context guarantees that the Home Safety Sound Detection System has the know-how to identify and address possible safety hazards.

3.2.4 Objective Formulation

Convolutional Neural Networks (CNNs) are being used in the Home Safety Sound Detection System with three distinct objectives. First, by utilizing the CNN algorithm's capabilities, the system seeks to do an analysis of the components suggestive of possible risks to home safety. This entails locating and examining particular sound patterns linked to safety issues, such glass breaking, sirens, or strange noises.

The second goal is to actually put the CNN algorithm to use in the development of a reliable sound detection system for home safety. This calls for the creation and application of an advanced model that can precisely categorize a range of auditory occurrences in real time. Finally, the goal also includes assessing the effectiveness of the home safety system, with a particular emphasis on the CNN algorithm's precision and efficiency.

The system's capacity to accurately recognize and react to sounds associated to safety will be evaluated through meticulous testing and validation procedures, guaranteeing its dependability and efficiency in augmenting home security.

3.2.5 Data Collection

A dataset comprising 500 audio samples featuring a variety of noises from people, machines, and vehicles has been sourced from Kaggle. This dataset will be used for data collection in the development of the Home Safety Sound Detection System. This dataset is a useful tool for testing and training the system's Convolutional Neural Networks (CNNs) algorithm. The collection includes a broad range of aural experiences, such as sounds made by machines, and people.

The Home Safety Sound Detection System is exposed to a wide range of auditory patterns thanks to the dataset's diversity, which also helps the CNN algorithm learn and generalize. The acquisition of this dataset was made possible by Kaggle, a reliable source of datasets and machine learning tools. This has enhanced the sound detection model's resilience and adaptability in identifying possible safety risks in a domestic setting. Figure 3.1 shows the dataset for the train model.

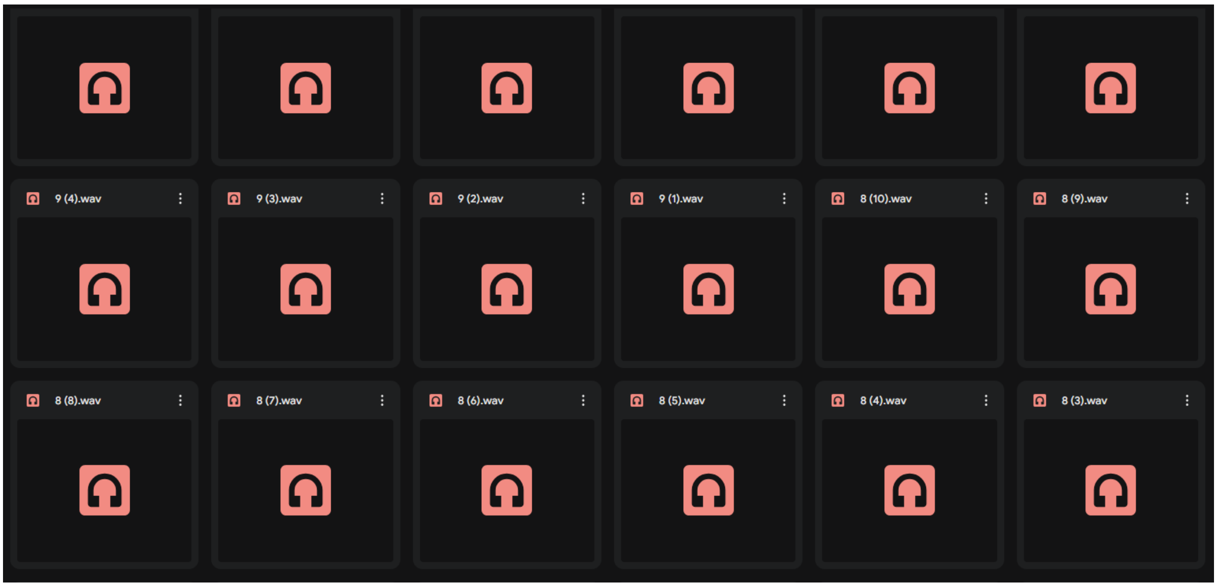


Figure 3. 1 Example of Dataset.

3.3 Design Phase

3.3.1 System Architecture

A complex system's high-level organization and structure, which describes its constituent parts, how they interact, and the design concepts that inform those parts, is referred to as system architecture. It acts as a construction plan for the system, offering a structure that guarantees the different components function together harmoniously to accomplish the system's objectives. Figure 3.1 below shows the system architecture for Home Safety Sound Detection.

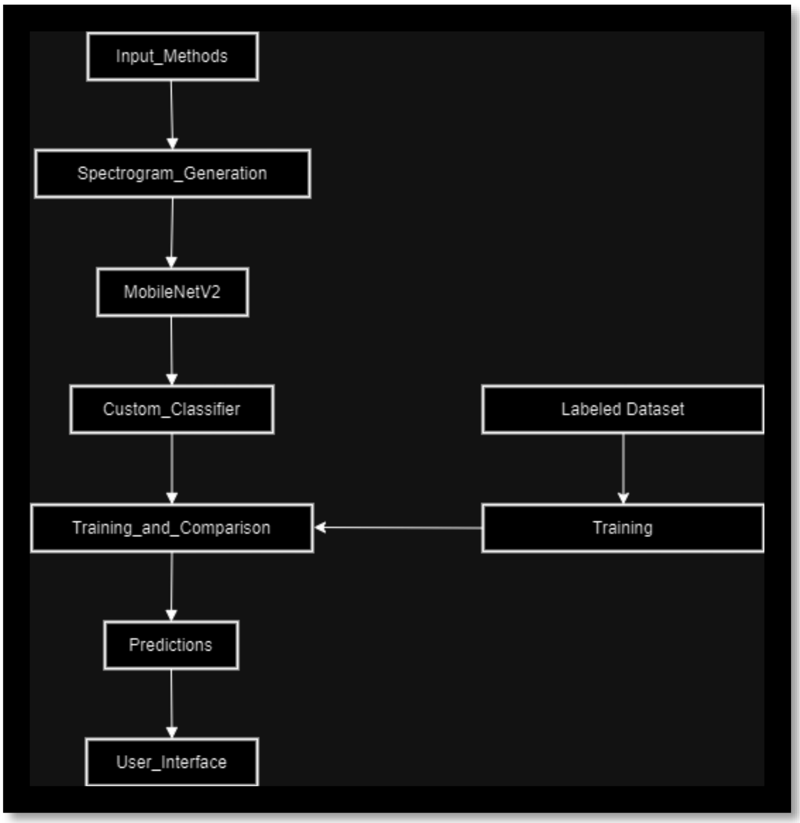


Figure 3. 2 System Architecture Home Safety Sound Detection.

First, the system provides users with many ways to interact with the application through input. Users can upload one.wav audio file at a time in Single File Mode and multiple.wav files at once in Multiple Files Mode. Users can record audio in real-time for up to 10 seconds by using the real-time recording mode. After receiving audio inputs, the system uses the Librosa library to build spectrograms. To improve visualization and facilitate feature extraction, audio data are first converted into Mel-frequency spectrograms and subsequently into decibel units.

MobileNetV2, a pre-trained Convolutional Neural Network (CNN) that was first trained on the ImageNet dataset, is the brains behind the system. Transfer learning is used to modify MobileNetV2 for sound classification tasks, with a focus on high-level feature extraction from spectrogram images. A custom classifier is formed by integrating additional thick layers on top of MobileNetV2. Activation functions such as ReLU are applied to intermediate layers and softmax to the final layer, flattening and reducing feature dimensions and generating a multi-class probability distribution.

In order for the system to compare incoming spectrogram features with learnt patterns from the training phase, it loads pre-trained weights from cnn_model_02.h5. The process of comparison is essential to producing precise forecasts. Preprocessing is applied to spectrogram images in order to align them with training circumstances and make sure they meet the model's expectations. After that, MobileNetV2 and the custom classifier layers process the acquired features to produce probability scores for every sound class.

Users can view the predictions generated by the system's inference process immediately using the Streamlit interface. Streamlit provides an interactive, user-friendly interface by displaying spectrograms, enabling audio playback, and providing prediction findings in an understandable way. Users can study spectrograms and related predictions by navigating through several pages created for each input method.

3.3.2 System Flowchart

The Home Safety Sound Detection System recognizes and reacts to any safety-related sounds in a domestic environment using a methodical flowchart. The first step in the procedure is for the user to choose a file mode, which controls how the application will analyze audio files. The user is invited to provide an audio file for analysis after choosing the mode. This user action sets off a sequence of methodical systemic actions. First, the user-provided audio file loads and is displayed by the system. In addition to ensuring transparency, this phase enables the user to confirm that the right file is being processed. Subsequently, the system transforms the audio data into a spectrogram, which is a picture that shows the audio signal's frequencies over time. For additional study, this spectrogram offers a comprehensive perspective of the frequency content of the audio.

The user can visually examine the frequency characteristics of the audio file by viewing the spectrogram that the system has developed. In order to get the spectrogram image ready for input into a neural network model, the system preprocesses it simultaneously. To meet the input requirements of the machine learning model that will be used for sound classification later on, this preprocessing phase optimizes the image. From the preprocessed spectrogram image, the system extracts relevant features using a pre-trained MobileNetV2 model. MobileNetV2, which is well-known for its efficacy and efficiency in the extraction of picture features, analyzes the spectrogram to extract pertinent patterns and information that are essential for sound classification tasks.

Using the features that were collected, the system trains a custom classifier on top of MobileNetV2 to predict the likelihood that different sound classes would appear in the audio. These courses usually include environmental situations (such storms or gunfire), particular events (like cries or engine noises), and background noise kinds (like ambient sounds). Lastly, the user is shown by the system the expected probability for each sound class. This output enables users to comprehend the sound profile and pinpoint any particular events or conditions recorded within the audio data. It also offers actionable insights into the composition of the studied audio file.

Figure 3.3 below shows the flowchart of system.

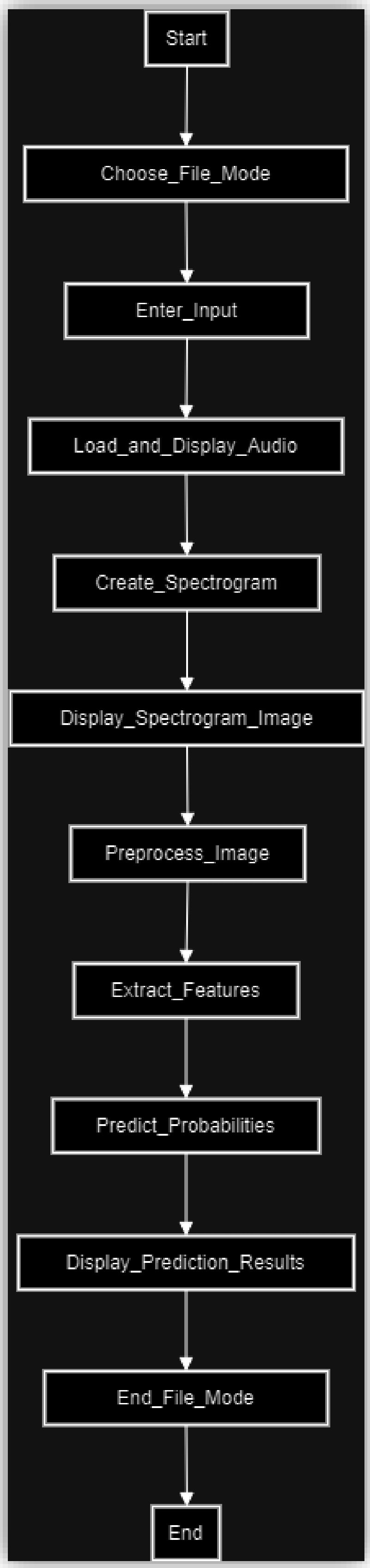


Figure 3. 3 Flowchart.

Convolutional Neural Networks are used in this flowchart to provide an organized method of sound detection, allowing for precise classification and prompt reactions to possible safety hazards.

3.3.3 User Interface Design

The process of developing the interactive and visual components of a software program or website that people interact with is known as user interface (UI) design. By creating user-friendly and aesthetically pleasing interfaces, it focuses on improving the user experience. In order to make an interface that is simple to use and navigate, user interface designers take into account layout, colours, typography, and interactive components. The aim is for users can interact with the system, understand the information presented and able to perform task efficiently.

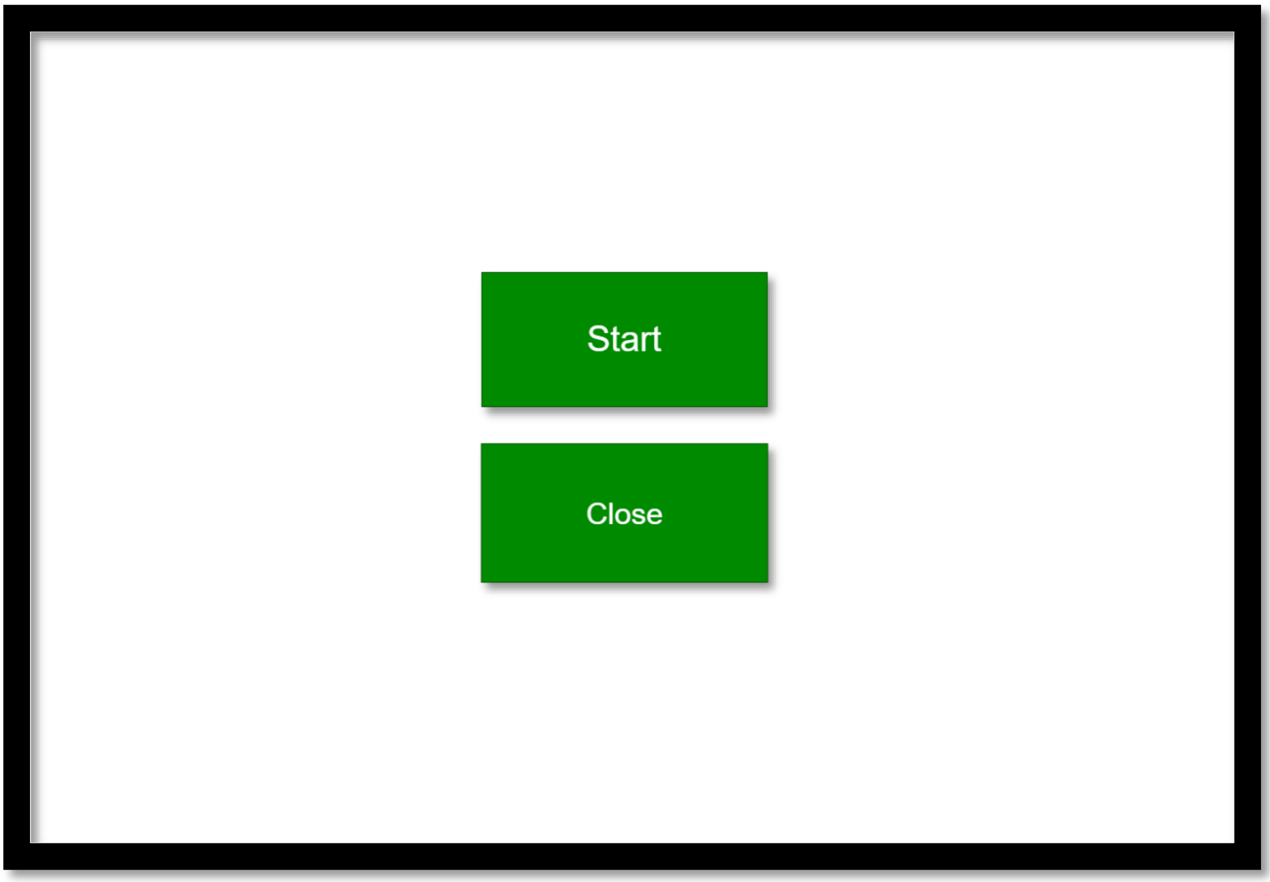


Figure 3. 4 First Page

Figure 3.3 is the main page for the system that contain with 2 buttons, which the button will be the start where after clicking the button, user will be directed to next page. While the second button will be close button where its function only to close the system.

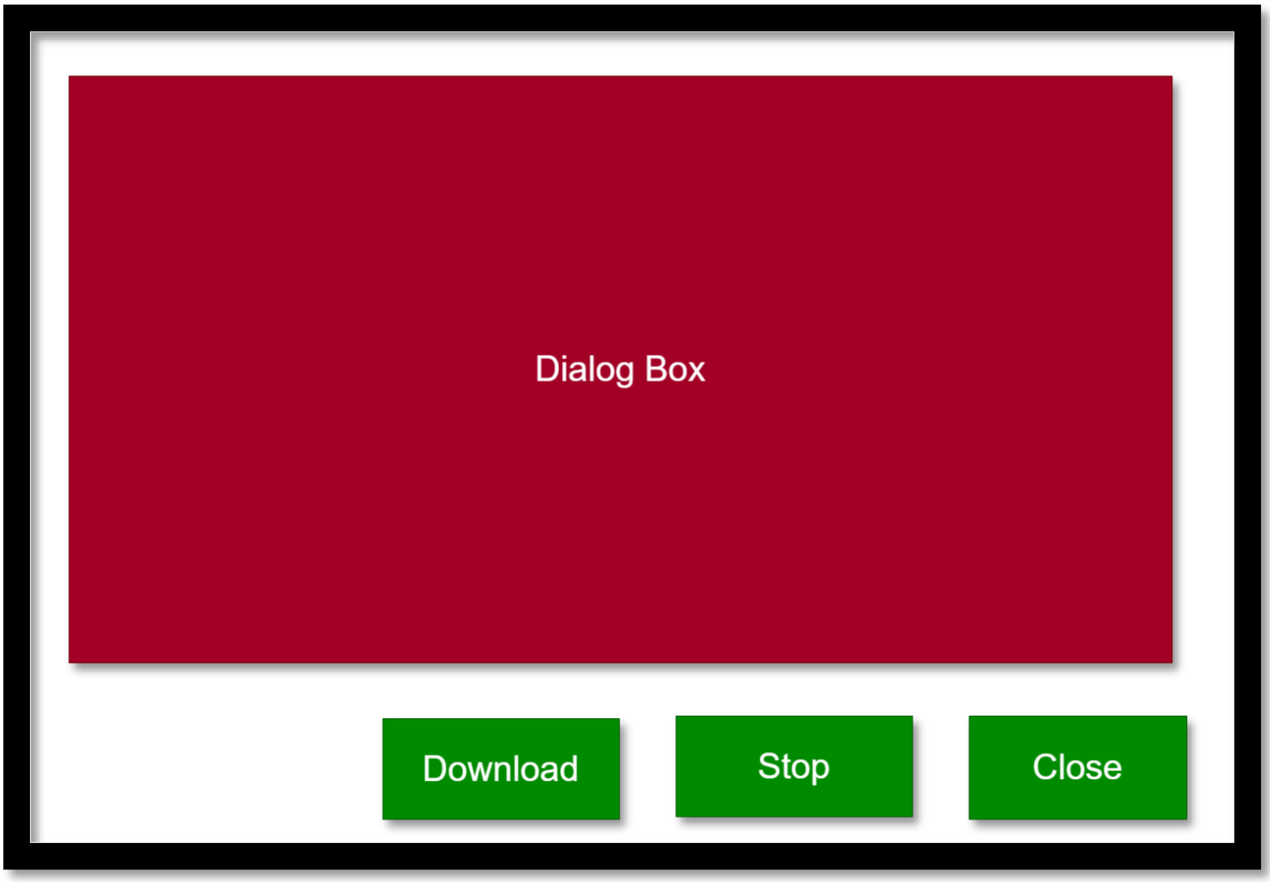


Figure 3. 5 Second Page

Figure 3.4 is the second page where consist of a big dialog box where it will report any detection that pick up by the system and 3 buttons where the first button is stop button where its main function is to stop the system from picking up more input. Next button is close button where the function is the same as the main page close button. While the last button is downloading button for user to download the content of the dialog box for further use.

Alert!!

Figure 3. 6 Alert Pop-up Notification

Figure 3.5 is the pop-up notification or alert notification, where it will pop-up to inform user for any activity.

3.4 Prototype Implementation

The implementation step entails converting a system's design into functional hardware and software components. This phase outlines the programming languages, frameworks, and tools required for system development in terms of software requirements. The system's scalability, interoperability with other technologies, and functionality all influence the software component selection. Regarding hardware, the implementation phase describes the infrastructure and physical parts required to run the system efficiently. This comprises any unique hardware needed for particular functionality, as well as servers, processors, memory, and storage.

3.4.1 Software Recommendation

The following software suggestions are recommended for the creation and documentation of the Home Safety Sound Detection System:

1. GoogleCollab for Model Training:
   Justification: Because it gives users access to free GPUs and TPUs, Google Colab is an effective tool for training machine learning models and allows for faster model training. It facilitates team collaboration on coding projects and makes data sharing and storage simple thanks to its interface with Google Drive.
   Use: Models may be trained using Google Colab, making machine learning research and development effective and scalable.

2. VSCode of User interface:
   Justification: Visual Studio Code (VSCode) is a flexible, lightweight code editor that offers outstanding assistance for Python development. It is ideally suited for creating and overseeing the user interface of the Home Safety Sound Detection System because of its extensible interface, built-in Git capabilities, and customizable interface.
   Use: VSCode can be used to design, code, and test the system's user interface components.

3. Draw.io for Drawing Flowcharts and System Architecture:

   Justification: Draw.io is an easy-to-use online diagramming tool that makes it possible to create system architectural diagrams and flowcharts. It is appropriate for graphically depicting the architecture and flow of the sound detection system because of its user-friendly interface and collaborative features.

   Use: draw.io may be used to create system architecture diagrams that describe the high-level organization of the Home Safety Sound Detection System as well as flowcharts that depict the successive phases in sound detection.

4. Microsoft Word for Documentation:

   Justification: Microsoft Word is a popular word processing program that offers a comfortable setting for in-depth documentation. System specs, user manuals, and project reports may all be created with it because to its text formatting, collaborative editing, and ability to include photos and diagrams.

   Use: The Home Safety Sound Detection System's functions, development process, and any other pertinent information can all be documented using Microsoft Word.

This software stack offers a well-rounded toolbox to effectively build, illustrate, and document the Home Safety Sound Detection System by integrating Jupyter for coding, VSCode for the user interface, draw.io for visual representations, and Microsoft Word for thorough documentation.

3.4.2 Hardware Recommendation

The following hardware recommendation is proposed for the Home Safety Sound Detection System development:

1. Laptop Model: Acer Nitro 5

   Justification: With specs appropriate for machine learning and software development, the Acer Nitro 5 is a versatile laptop. The system's dedicated graphics card, gaming-focused features, and strong performance enable it to manage the computational demands of developing sound detection systems.
2. Processor: Intel® Core™ i5-8300 CPU @ 2.50 GHz

Justification: The Intel Core i5-8300 processor offers a performance and power efficiency balance. Its numerous cores and threads enable parallel processing and multitasking, which are crucial for machine learning activities and the seamless operation of the development environment.

3. RAM: 8GB
   Justification: Moderately sized software development and machine learning projects can benefit from an 8GB RAM configuration. Without major bottlenecks, it enables the efficient handling of code, data, and model training procedures.
4. Operating System: Windows 10
   Justification: Windows 10 is a popular operating system that works well with software development. It offers a simple user interface, robust software support, and interoperability with other frameworks and development tools.
5. System Type: 64-bit operating system, x64-based processor
   Justification: By utilizing the laptop's hardware to its fullest extent, a 64-bit operating system makes it possible to do complicated computations and enormous datasets quickly. The CPU architecture, which is based on x64, guarantees compatibility with contemporary software and development tools.

The Acer Nitro 5 laptop, Intel Core i5 CPU, 8GB RAM, Windows 10 operating system, and 64-bit architecture that make up this hardware configuration offer a strong platform for the creation and operation of the Home Safety Sound Detection System. Performance, cost, and compatibility are all balanced for efficient system testing and development.

3.5 Evaluation Phase

An important stage in the creation of the Home Safety Sound Detection System is the evaluation phase, which evaluates the efficacy and performance of the system. Examining the system's accuracy and efficiency two critical components that together define its overall effectiveness is part of this comprehensive review.

The evaluation phase provides a comprehensive understanding of the capabilities of the Home Safety Sound Detection System by synthesizing the findings from accuracy and efficiency tests. The process of iterative optimization involves careful consideration of trade-offs between accuracy and efficiency. To establish a balance that meets project objectives, model refinement, hyperparameter adjustments, and/or exploration of alternative architectures may be necessary.

3.5.1 Accuracy

Accuracy, precision, recall, and F1 score are among the important classification metrics used to assess the accuracy of the Home Safety Sound Detection System. When assessing the model's effectiveness in sound event detection tasks, several indicators are crucial. An overall measure of correctness is provided by the accuracy, which is computed as the ratio of correctly classified examples to the total instances. The precision, defined as the ratio of true positive predictions to the sum of true positives and false positives, emphasizes the accuracy of positive forecasts.

Conversely, recall, sometimes referred to as sensitivity, is a measure of how well the model captures all real positive cases; it is computed as the ratio of genuine positives to the total of false negatives and true positives. The F1 score is a balanced statistic that considers both false positives and false negatives. It is calculated as the harmonic mean of precision and recall. All these indicators add up to a thorough evaluation of the system's accuracy, providing information about how well it can detect sound events. The following are the formulas for these metrics:

Accuracy:

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

Precision:

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

Recall (Sensitivity):

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

F1 Score:

$$F1 = \frac{2.(Precision.Recall)}{Precision + Recall}$$

True positives are indicated by TP, true negatives by TN, false positives by FP, and false negatives by FN in this case. The combination of these criteria provides a detailed assessment that makes it easier to comprehend how accurate the Home Safety Sound Detection System is at identifying pertinent sound events.

3.5.2 Efficiency

Throughout the training and inference stages, the Home Safety Sound Detection System's effectiveness is evaluated using a variety of metrics. Training efficiency is measured by timing how long the model takes to train on the dataset and keeping an eye on how much CPU and GPU are being used. In order to maximize training times and resource usage, batch processing algorithms are investigated. During the inference stage, the effectiveness of the system is evaluated based on how quickly reliable event predictions are made and how well resources are used in batch or real-time processing.

To increase computational efficiency, strategies including hardware acceleration, quantization, and model pruning may be taken into consideration. Real-time testing scenarios are applied to the system in order to evaluate its practical usability and responsiveness. User feedback is collected to capture subjective opinions of system efficiency. A thorough grasp of how successfully the Home Safety Sound Detection System strikes a balance between processing demands and precise sound event detection is ensured by the documentation of these efficiency assessments and optimization suggestions. In order to improve accuracy and efficiency, the model and parameters are refined iteratively during the optimization process.

3.6 Documentation

One essential element that records all of the in-depth information about the design, operation, and deployment of the Home Safety Sound Detection System is its documentation. This documentation offers insights into the architecture, operation, and assessment of the system, making it an invaluable tool for developers, users, and stakeholders. The intention is to provide a resource that guarantees data security, logs errors for future correction, and adheres to moral standards to prevent plagiarism and original work.

3.6.1 Gantt Chart

Table 3. 2 Gantt Chart

ACTIVITIES	SEMERTER 5														SEMESTER 6													
	DURATION (WEEK)																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	1	2	3	4	5	6	7	8	9	10	11	12	13	14
PRELIMINARY PHASE	■	■	■	■	■																							
Proposal	▨	▨	▨	▨	▨																							
Discuss with Supervisor		▨	▨	▨																								
Data Collection			▨	▨	▨																							
Literature Review			▨	▨	▨																							
ANALYSIS PHASE						■	■	■	■	■																		
Study the algorithm involved						▨	▨	▨																				
Choose the best technique									▨	▨																		
DESIGN PHASE											■	■	■	■														
Design the proposed system											▨	▨	▨	▨														
DEVELOPMENT PHASE															■	■	■	■	■	■	■	■	■	■				
Implement and develop system															▨	▨	▨	▨	▨	▨	▨	▨	▨	▨				
EVALUATION PHASE																									■	■	■	
Test and evaluate the system with accuracy																									▨	▨	▨	
DOCUMENTATION	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Full Report	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨

3.7 Conclusion

In summary, a methodical and well-structured research technique was used in the development of the Home Safety Sound Detection System employing Convolutional Neural Networks (CNNs). A comprehensive issue statement formulation, literature review, and knowledge acquisition were all part of the preparatory phase, which gave rise to a strong basis for comprehending the difficulties and current solutions in sound event detection. The goals were well-defined and included researching possible risks to home security, putting in place a CNN-based sound detection system, and assessing the effectiveness and accuracy of the system.

In the following stage of data collection and preparation, 500 audio samples pertaining to different sound categories associated with human, and machine activities were acquired. During the design phase, the system architecture, system flowchart, and user interface design were developed with the aim of optimizing CNNs' capabilities in sound event detection. The logistic regression pseudocode offered a detailed implementation road map.

As for the prototype implementation, the technique suggested using certain software tools including Microsoft Word for documentation, draw.io for flowchart design, VSCode for the user interface, and GoggleCollab for Model Training. An Acer Nitro 5 laptop with an Intel® CoreTM i5-8300 CPU, 8GB RAM, and Windows 10 operating system was suggested as hardware.

Accuracy and efficiency were the two main areas of attention for the system evaluation. Precision, recall, and F1 score were computed as accuracy measures, and training and inference times, resource usage, and real-time testing scenarios were used to gauge efficiency. The system's architecture, algorithms, training procedure, assessment measures, and optimization techniques were all covered in detail throughout the documentation phase, which served as a process summary. It added to the iterative aspect of the study approach by offering suggestions for more development.

The research technique, which leverages CNNs to handle the issues associated with detecting potential dangers in home surroundings, essentially defined a thorough and iterative framework for the creation and evaluation of the Home Safety Sound Detection System. From problem formulation to documentation, the methodical approach guarantees repeatability, transparency, and lays the groundwork for future developments in the field of home safety sound detection.

CHAPTER 4

RESULT AND FINDING

The outcomes and conclusions of the study are presented in this chapter, with particular attention paid to the assessment of the classification system and model. It covers the conceptual framework, project interfaces, program codes, and assessment techniques in great depth.

4.1 Conceptual Framework

A conceptual framework is a theoretical structure that delineates the essential details and connections among all the project's processes. It begins with the user's input and ends with the output that is shown to the user again.

1. Input:

Input (User Interface - Streamlit): Users upload audio files by interacting with the Streamlit online interface. Users can upload one or more audio files using the file uploader widget (st.file_uploader) offered by Streamlit. Streamlit is adaptable to a range of user requirements since it enables users to record audio straight from the web interface for real-time input.

2. Processing Stage:

The Librosa library is used to process the audio files after they are uploaded. After reading the audio file, Librosa's librosa.load() function returns the sampling rate (sr) and the audio time series (y). This step is essential because it converts the unprocessed audio file into a format that can be used for additional processing and analysis.

3. Preprocessing Stage:

Librosa's librosa.feature.melspectrogram() is then used to transform the audio signal into a spectrogram, which is a graphic depiction of the frequency content of the audio signal over time. With librosa.power_to_db(), the mel spectrogram is transformed into a decibel (log) scale. The spectrogram is shown using librosa.display.specshow() and is visualized using Matplotlib. The spectrogram is then saved as an image file, ready for the next step of picture-based classification.

4. Feature Extraction Stage:

This step involves loading the stored spectrogram image using Keras's image.load_img() method and resizing it to 224 by 224 pixels in order to comply with the model's input specifications. Image.img_to_array() is used to convert the image to a NumPy array, then np.expand_dims() is used to expand the image to include a batch dimension. Next, the preprocess_input() function of MobileNetV2 is used to preprocess the image array in accordance with the preprocessing procedures used during the model's training. After the spectrogram image has been preprocessed, it is loaded without the top classification layers (include_top=False) and pretrained on the ImageNet dataset using the MobileNetV2 base model. As a result, the image's high-level features are represented in a feature map.

5. Prediction Stage:

For classification, additional custom dense layers are applied on top of the MobileNetV2 base model. This comprises an output layer with 5 neurons (one for each class: background, scream, engine, storm, and gunshot) with softmax activation for outputting classification probabilities, a dense layer with 1024 neurons and ReLU activation, and a flatten layer to convert the feature map into a 1-dimensional vector. To make sure the model is optimized for audio signal classification, it is trained on the spectrogram images that match to the audio classes.

6. Output Stage:

To derive classification probabilities, the custom dense layers are applied to the extracted features from the MobileNetV2 base model. The spectrogram image and the classification probabilities for each predefined category are then shown by Streamlit. The findings display the likelihood of each class for the submitted audio file in an easy-to-use format. This blend of deep learning, web-based user interaction, and audio signal processing guarantees precise categorization of audio signals and user-friendly result presentation.

The conceptual structure for the categorization model is displayed in Figure 4.1.

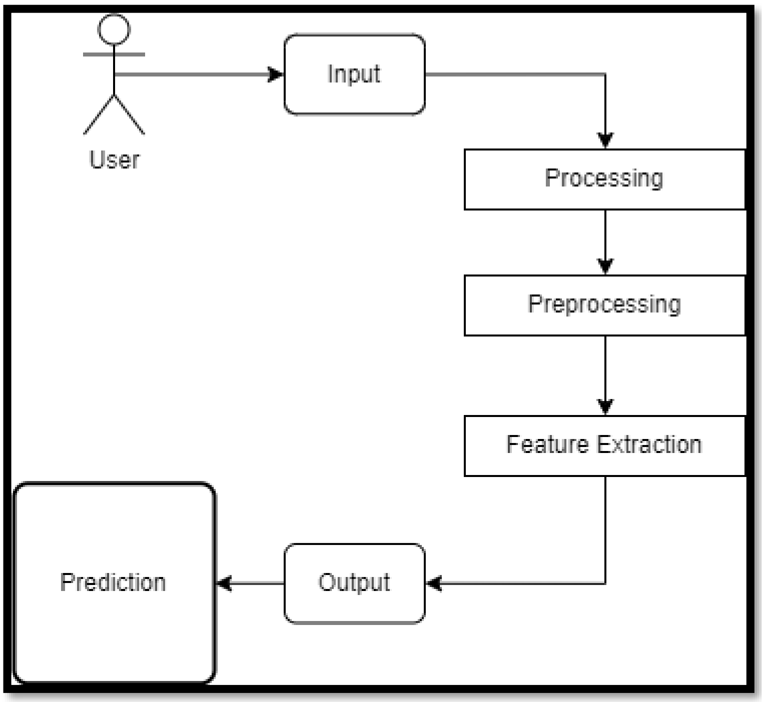


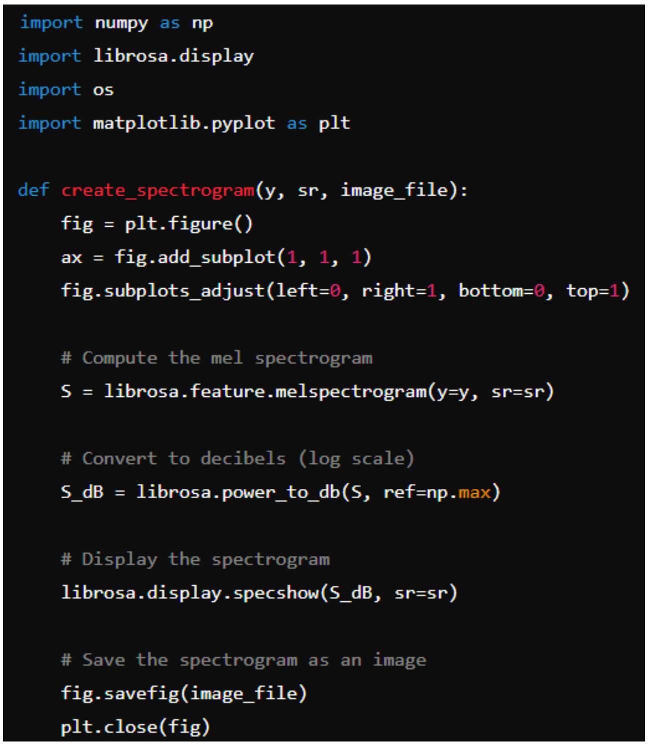
Figure 4. 1 Conceptual Framework

4.2 Program Codes for Algorithm

In this section, we outline the process of converting audio files into spectrograms and then using the MobileNetV2 architecture to train and evaluate a deep learning model for sound classification. This approach makes use of convolutional neural networks (CNNs) and transfer learning.

4.2.1 Audio to Spectrogram Transformation

The project's first step is to create spectrogram images from raw audio recordings. Spectrograms are graphic depictions of a sound signal's frequency spectrum as it changes over time. This conversion is essential because it makes it possible to classify audio using computer vision and image processing methods. Below is the code for this function.



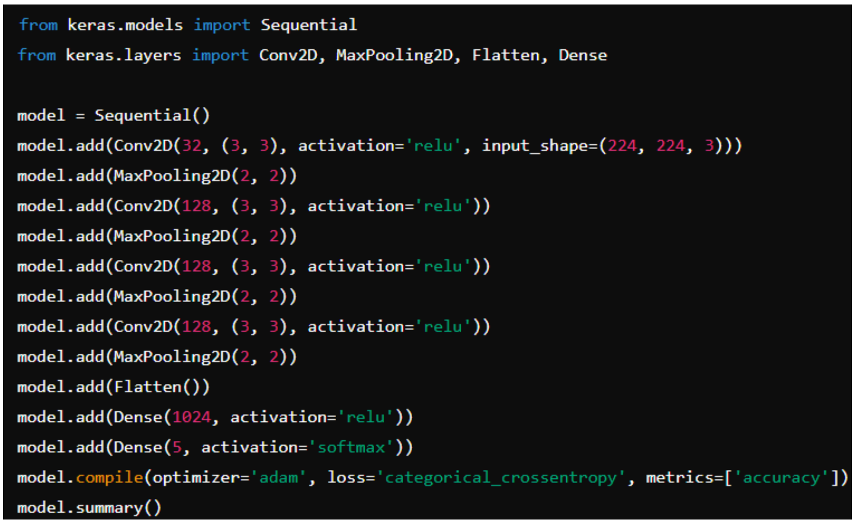
```
import numpy as np
import librosa.display
import os
import matplotlib.pyplot as plt

def create_spectrogram(y, sr, image_file):
    fig = plt.figure()
    ax = fig.add_subplot(1, 1, 1)
    fig.subplots_adjust(left=0, right=1, bottom=0, top=1)

    # Compute the mel spectrogram
    S = librosa.feature.melspectrogram(y=y, sr=sr)

    # Convert to decibels (log scale)
    S_dB = librosa.power_to_db(S, ref=np.max)

    # Display the spectrogram
    librosa.display.specshow(S_dB, sr=sr)

    # Save the spectrogram as an image
    fig.savefig(image_file)
    plt.close(fig)
```

Figure 4. 2 create_spectogram code.

Mel spectrograms, which offer a time-frequency representation of the audio signal, are produced by the 'create_spectrogram' function using the 'librosa' library. The power spectrogram can be converted to decibels to improve the data's visual representation. After being exported as PNG pictures, the spectrograms can be utilized as input data for image classification algorithms.

4.2.2 Convolutional Neural Network (CNN) Layers

Utilizing a Convolutional Neural Network (CNN), the spectrogram images are processed. Because convolutional layers allow CNNs to record spatial hierarchies, they are especially well-suited for image data.



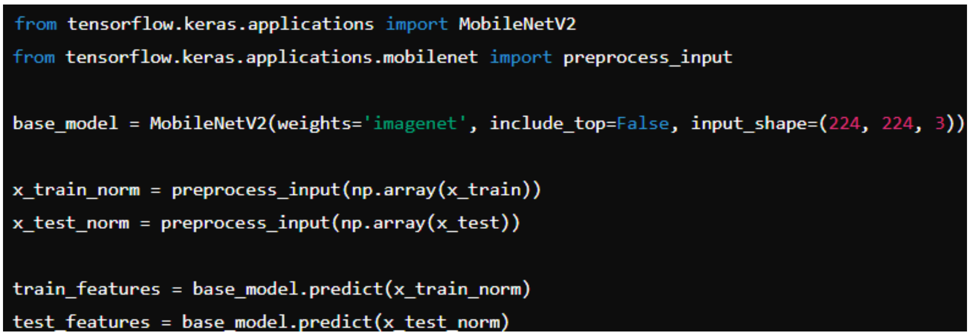
```
from keras.models import Sequential
from keras.layers import Conv2D, MaxPooling2D, Flatten, Dense

model = Sequential()
model.add(Conv2D(32, (3, 3), activation='relu', input_shape=(224, 224, 3)))
model.add(MaxPooling2D(2, 2))
model.add(Conv2D(128, (3, 3), activation='relu'))
model.add(MaxPooling2D(2, 2))
model.add(Conv2D(128, (3, 3), activation='relu'))
model.add(MaxPooling2D(2, 2))
model.add(Conv2D(128, (3, 3), activation='relu'))
model.add(MaxPooling2D(2, 2))
model.add(Flatten())
model.add(Dense(1024, activation='relu'))
model.add(Dense(5, activation='softmax'))
model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
model.summary()
```

Figure 4. 3 CNN layers code

Multiple convolutional and max-pooling layers make up the CNN model, which is in charge of extracting features from the input images. Convolution operations are applied through filters by the ‘Conv2D’ layers to identify features including textures, edges, and patterns. By lowering the spatial dimensions of the feature maps, the ‘MaxPooling2D’ layers downsample the data and lower computing complexity. Lastly, the classification is carried out by the dense layers which include a softmax layer using the features that were extracted.

4.2.3 MobileNetV2 Function

The spectrogram images are processed through feature extraction using MobileNetV2. The efficiency and strong performance of this pre-trained model are well-known in embedded and mobile vision applications.



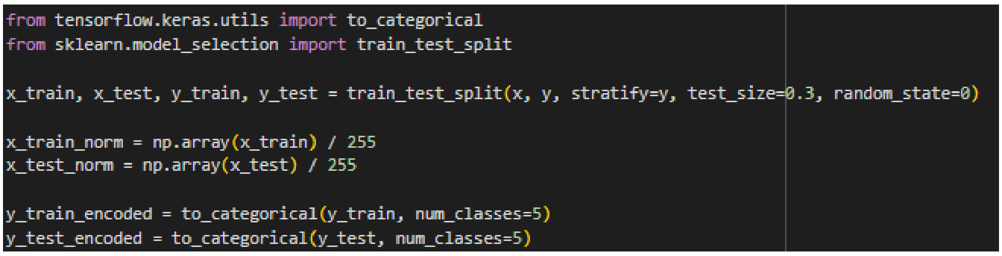
```
from tensorflow.keras.applications import MobileNetV2
from tensorflow.keras.applications.mobilenet import preprocess_input

base_model = MobileNetV2(weights='imagenet', include_top=False, input_shape=(224, 224, 3))

x_train_norm = preprocess_input(np.array(x_train))
x_test_norm = preprocess_input(np.array(x_test))

train_features = base_model.predict(x_train_norm)
test_features = base_model.predict(x_test_norm)
```

Figure 4. 4 MobileNetV2 code.

For feature extraction, MobileNetV2 is a lightweight, effective deep learning model. We make use of transfer learning by leveraging a pre-trained model, which makes the model stronger due to its extensive training on a big dataset (ImageNet). By giving our model solid feature representations, this greatly enhances its performance.

4.2.4 Training and Testing

A new model is trained on the extracted characteristics in order to tackle the classification job. The training and testing sets of the dataset are separated in order to assess the performance of the model.



```
from tensorflow.keras.utils import to_categorical
from sklearn.model_selection import train_test_split

x_train, x_test, y_train, y_test = train_test_split(x, y, stratify=y, test_size=0.3, random_state=0)

x_train_norm = np.array(x_train) / 255
x_test_norm = np.array(x_test) / 255

y_train_encoded = to_categorical(y_train, num_classes=5)
y_test_encoded = to_categorical(y_test, num_classes=5)
```

Figure 4. 5 Model Training Code

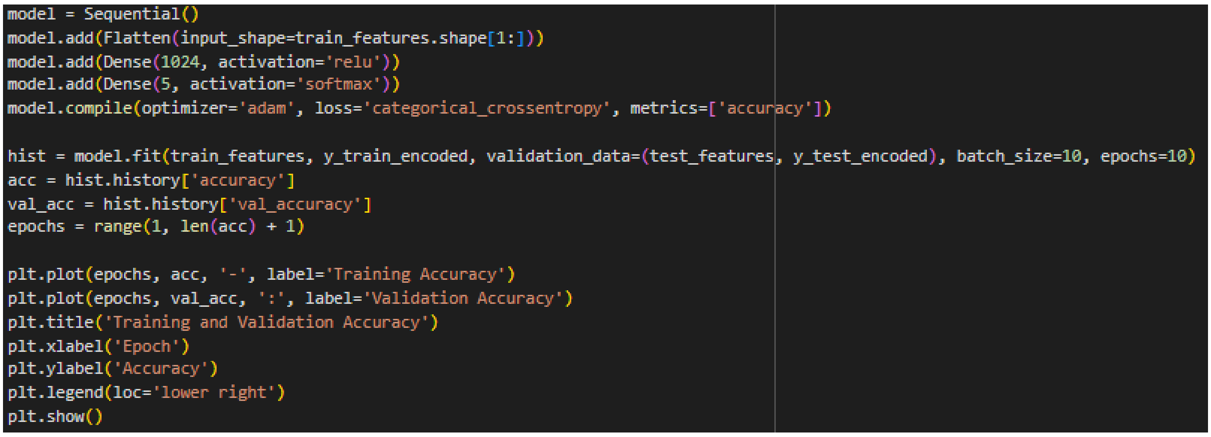
The dataset is split into training and testing sets using the train_test_split method from sklearn.model_selection, where x and y stand for the features (spectrogram images) and labels (sound classifications), respectively. For balanced class representation, the stratify=y option makes sure that the split keeps the same percentage of each class in both sets. Test_size=0.3 designates that 30% of the data be utilized for the testing set and the remaining 70% for training.

Table 4. 1 Percentage Split of Dataset

Split	Percentage Ratio	Dataset
Training	70%	350
Test	30%	150

There is no random state. To ensure reproducibility, a parameter regulates the random number generator that is used to shuffle the data. We then use tensorflow.keras.utils' to_categorical function to transform the labels into a one-hot encoded format. Through this procedure, binary vectors representing each class are used to represent categorical labels in a binary matrix format.

The argument num_classes=5 denotes that there are five distinct classes (background, scream, engine, storm, and gunshot, for example).



```
model = Sequential()
model.add(Flatten(input_shape=train_features.shape[1:]))
model.add(Dense(1024, activation='relu'))
model.add(Dense(5, activation='softmax'))
model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])

hist = model.fit(train_features, y_train_encoded, validation_data=(test_features, y_test_encoded), batch_size=10, epochs=10)
acc = hist.history['accuracy']
val_acc = hist.history['val_accuracy']
epochs = range(1, len(acc) + 1)

plt.plot(epochs, acc, '-', label='Training Accuracy')
plt.plot(epochs, val_acc, ':', label='Validation Accuracy')
plt.title('Training and Validation Accuracy')
plt.xlabel('Epoch')
plt.ylabel('Accuracy')
plt.legend(loc='lower right')
plt.show()
```

Figure 4. 6 Model Test Code

The Sequential API provided by Keras is used to define the model. The process begins with a flattened layer that creates a 1D vector from the 2D feature maps from MobileNetV2. The Dense layer, which functions as a fully connected layer to learn complicated representations, comes next. It has 1024 units and ReLU activation. With five units and a softmax activation function, the final layer is a dense layer that outputs the probabilities for each of the five classes.

1. Compilation:

The model is assembled using the Adam optimizer, a well-known tool for deep neural network training due to its efficiency and efficacy. For multi-class classification issues, the loss function categorical_crossentropy is employed. Accuracy is another statistic that the model is configured to track.

2. Training:

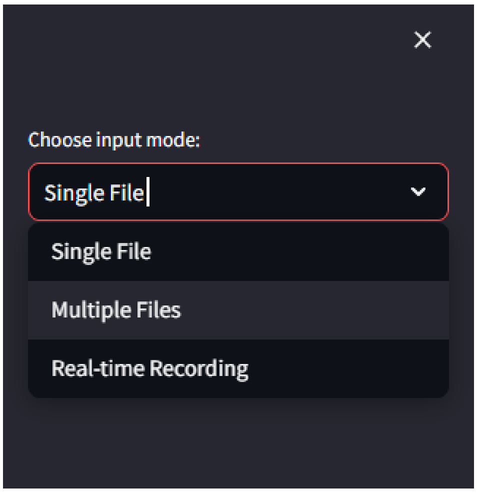
The extracted features (train_features) and the matching one-hot encoded labels (y_train_encoded) are used to train the model. At the conclusion of each epoch, the validation_data parameter instructs the model to assess its performance on the testing set (test_features and y_test_encoded). Ten epochs of training are carried out in a batch size of ten.

3. Plotting:

Hist.history['accuracy'] contains the training accuracy, while hist.history['val_accuracy'] contains the validation accuracy. To see how the model performed during training, these are then plotted over the epochs. This aids in determining if the model is overfitting or underfitting and how effectively it is learning.

4.3 Prototype Interface

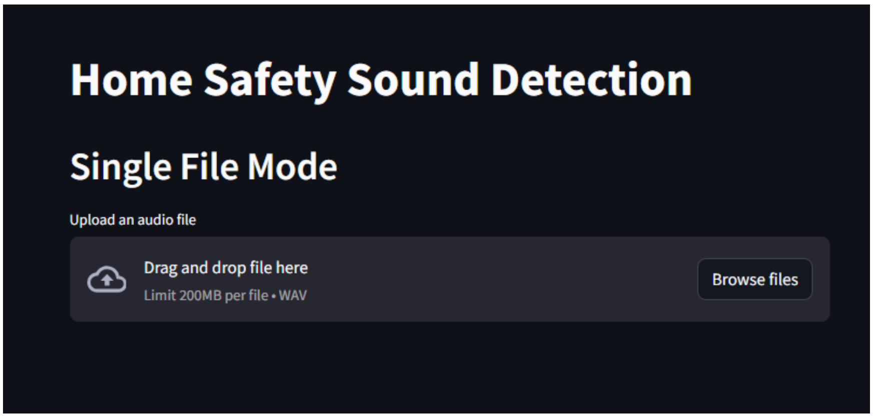
An essential part of the system that gives consumers a smooth and easy-to-use interaction experience is the user interface. Three separate pages, each intended to support a different functionality, make up this project. Users can move between the several pages of the basic interface using a dropdown menu, as shown in Figure 4.7.



Choose input mode: Single File

Figure 4. 7 Dropdown menu for each page

The input of a single audio file takes up the first page, as seen in Figure 4.8. For the system to process, users must submit one audio file at a time.



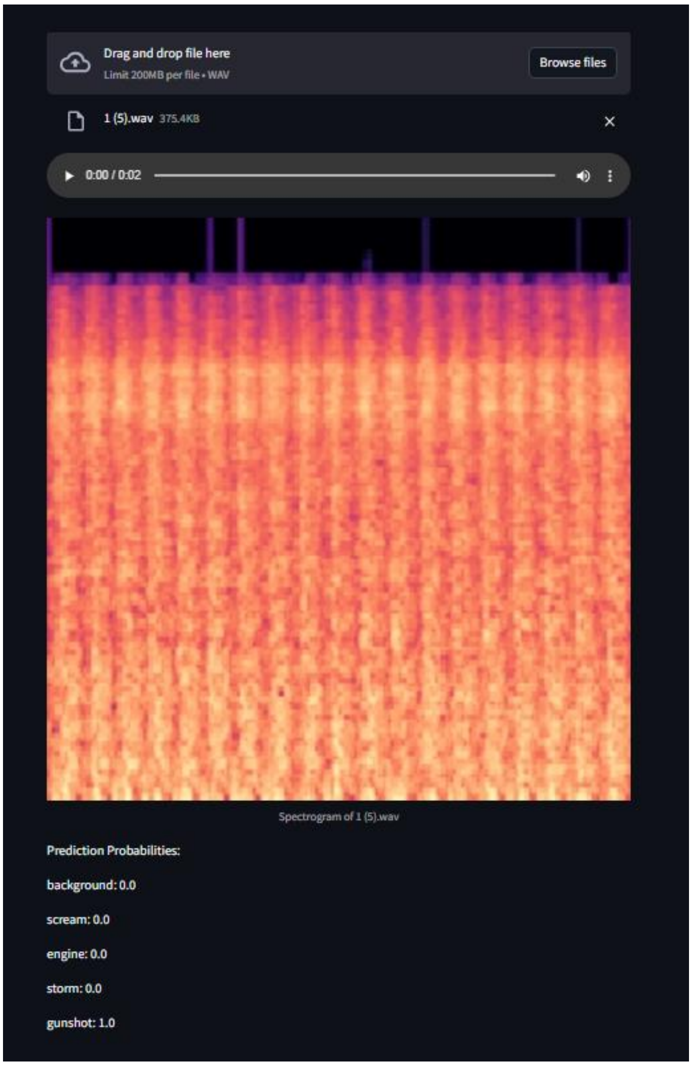
Home Safety Sound Detection

Single File Mode

Upload an audio file

Figure 4. 8 Single input user interface

Upon processing the input, the system displays the results as shown in Figure 4.9. The output includes a playback button enabling users to replay the submitted audio, a visual representation of the spectrogram, and the prediction probabilities generated by the system.



Prediction Probabilities:

background: 0.0

scream: 0.0

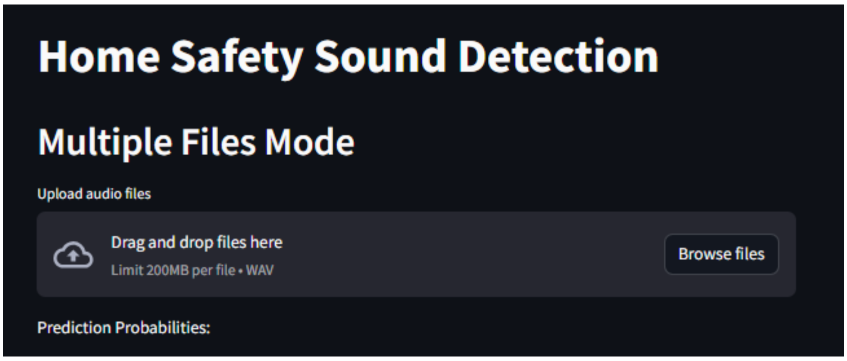
engine: 0.0

storm: 0.0

gunshot: 1.0

Figure 4. 9 Output single file

The second page of the interface is designed for multiple audio file inputs. As illustrated in Figure 4.10 below, this interface allows users to upload multiple audio files simultaneously for processing by the system.



Home Safety Sound Detection

Multiple Files Mode

Upload audio files

Prediction Probabilities:

Figure 4. 10 Multiple input user interface

The output for multiple audio files includes a playback button for each audio file, accompanied by its corresponding spectrogram image, and the prediction probabilities generated by the system. This process is depicted in Figures 4.11 and 4.12.

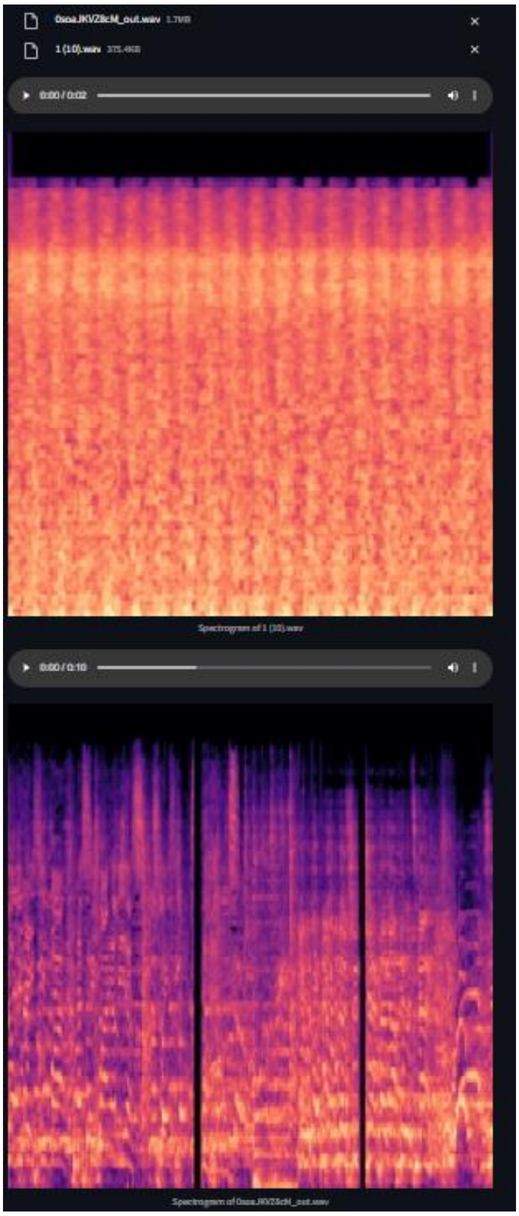
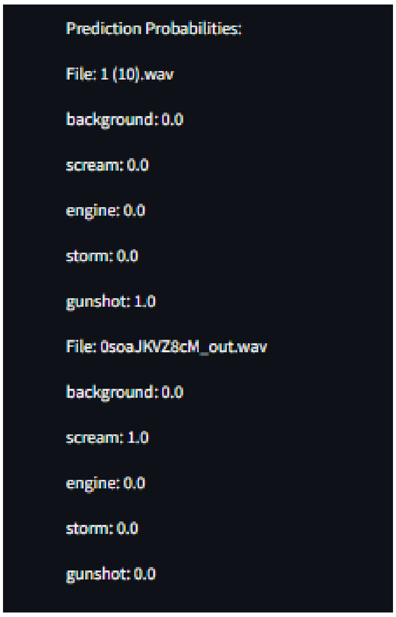

0soaJKVZ8cM_out.wav 1.7MB
1 (10).wav 375.4KB
0:00 / 0:02
Spectrogram of 1 (10).wav
0:00 / 0:10
Spectrogram of 0soaJKVZ8cM_out.wav

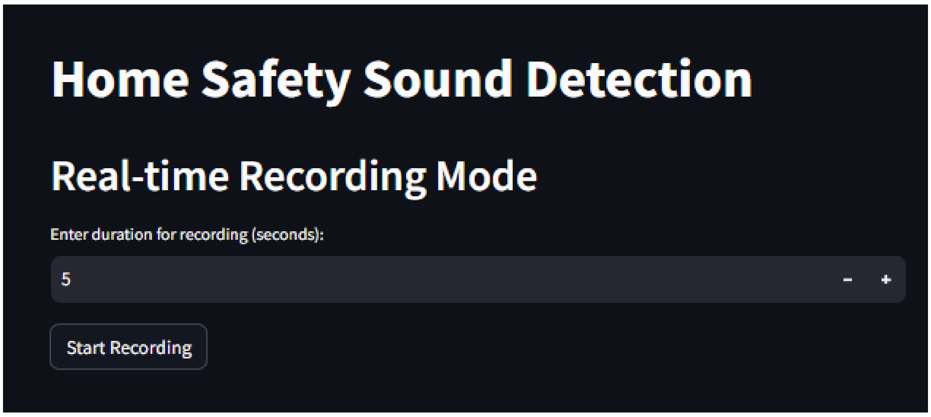
Figure 4. 11 Output multiple file



```
Prediction Probabilities:
File: 1 (10).wav
background: 0.0
scream: 0.0
engine: 0.0
storm: 0.0
gunshot: 1.0
File: 0soaJKVZ8cM_out.wav
background: 0.0
scream: 1.0
engine: 0.0
storm: 0.0
gunshot: 0.0
```

Figure 4. 12 Output multiple file

As seen in Figure 4.13, the last page is devoted to real-time recording. By selecting the "Start Recording" button on this page, users can record audio in real-time for a maximum of 10 seconds.



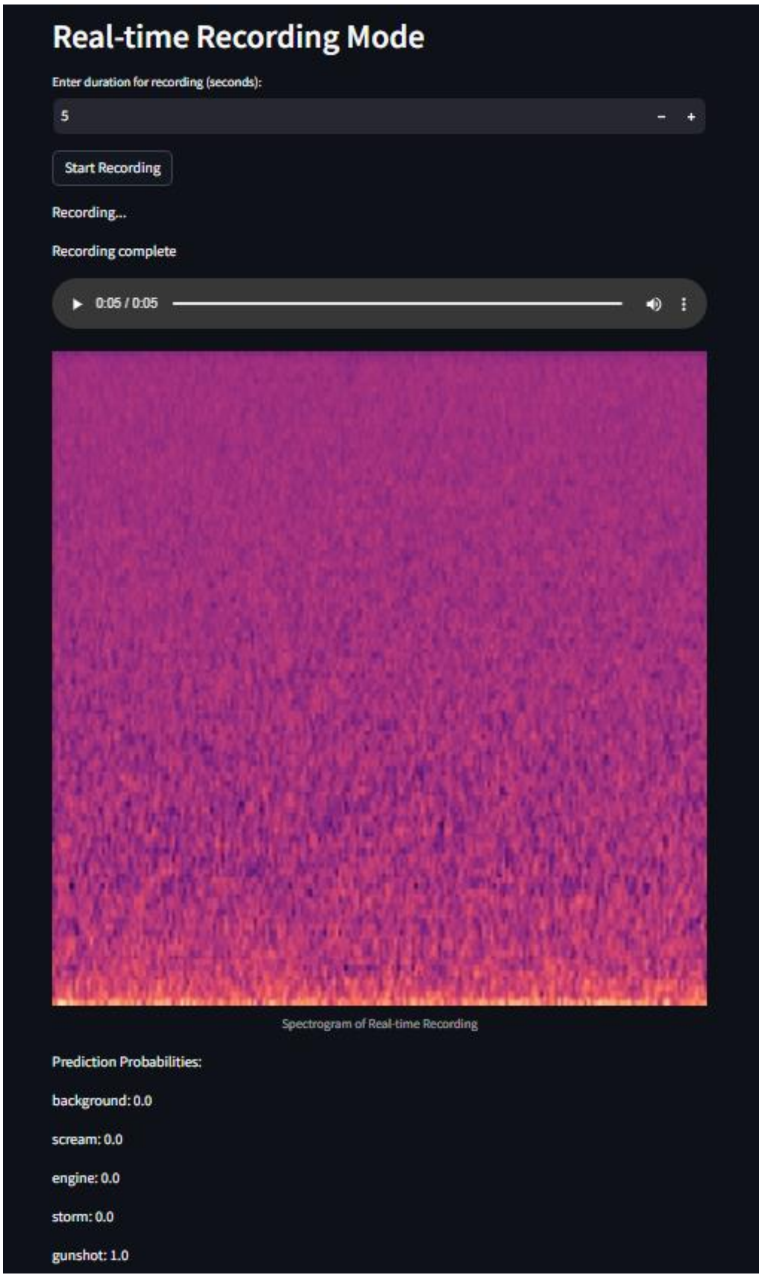
Home Safety Sound Detection

Real-time Recording Mode

Enter duration for recording (seconds): 5

Figure 4. 13 Real-time input user interface

The system uses the microphone to record any sound input when it is turned on. The system converts the recorded audio into a spectrogram for processing and prediction as soon as the recording stops.



Real-time Recording Mode

Enter duration for recording (seconds): 5

Recording...

Recording complete

Prediction Probabilities:

background: 0.0

scream: 0.0

engine: 0.0

storm: 0.0

gunshot: 1.0

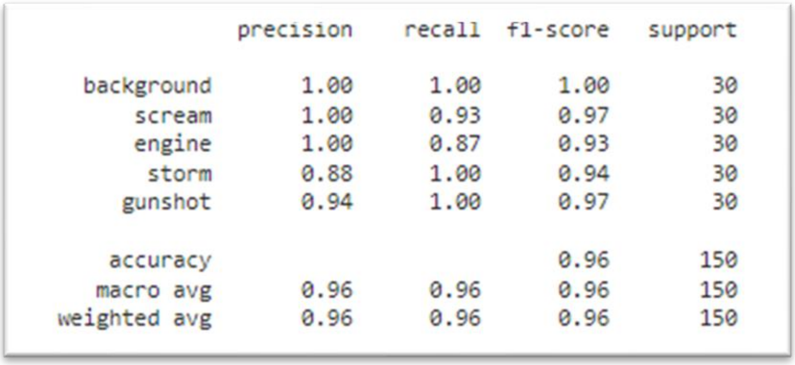
Figure 4. 14 Real-time output

As illustrated in Figure 4.14, the output comprises the prediction probabilities produced by the system, a spectrogram image of the audio, and a playing button for the recorded audio.

4.4 Evaluation Result

4.4.1 Classification Report

A machine learning model's performance on a classification job is thoroughly summarized in a classification report. It offers comprehensive measurements for every class in the dataset, enabling a thorough comprehension of the model's performance across several areas. Figure 4.15 shows the classification report taken from the model’s performance.



```
              precision    recall  f1-score   support

  background       1.00      1.00      1.00        30
      scream       1.00      0.93      0.97        30
      engine       1.00      0.87      0.93        30
       storm       0.88      1.00      0.94        30
     gunshot       0.94      1.00      0.97        30

    accuracy                           0.96       150
   macro avg       0.96      0.96      0.96       150
weighted avg       0.96      0.96      0.96       150
```

Figure 4. 15 Classification report.

4.4.1.1 Class-specific Metrics

The model received flawless results for the background class in every metric. The model correctly categorized all instances of background sounds with no false positives or false negatives, with precision, recall, and f1-score all at 1.00. This suggests that the model's ability to distinguish background noise is quite dependable.

With a score of 1.00 in the scream class, the model also showed excellent precision, indicating that all of the model's predictions about screams were accurate. On the other hand, the recall for this class was 0.93, meaning that 7% of the actual scream instances were missed by the model. The 0.97 f1-score indicates a good trade-off between recall and precision, but it also shows a small missing piece in the capture of all scream occurrences.

With a score of 1.00 in the engine class, the model demonstrated excellent precision performance, indicating that all projected engine noises were accurate. But the recall was 0.87, meaning 13% of real engine noises were missed by the model. As a result, the f1-score was

0.93, which still shows excellent performance but implies that there is potential for development in terms of identifying all engine noises.

The model's precision for the storm class was 0.88, which means that 12% of the storm sounds that were predicted were off. Still, with a recall of 1.00, it was able to accurately identify every real storm sound. Although the model is quite good at identifying storm noises, some predictions may be incorrectly assigned to this category, as indicated by the f1-score of 0.94.

With a precision of 0.94 and a recall of 1.00, the gunshot class demonstrated that the model was able to identify all gunshot sounds; nevertheless, it had a 6% error rate in its predictions, leading to some false positives. Gunshot sound recall and precision are well balanced, as evidenced by the f1-score of 0.97.

4.4.1.2 Overall Metrics

With an overall accuracy of 0.96, the model successfully identified 96% of the examples in all classes. This great accuracy is a result of the model's resilience and efficiency in differentiating between various sound kinds. The model's average performance across all classes, treating each equally, is represented by the macro average metrics, which have precision, recall, and f1-score all at 0.96. This indicates that the model is not biased against any certain class and consistently performs well across all sound categories.

The f1-score at 0.96, precision, and recall are also reported using the weighted average metrics. These averages show that the model continues to perform well even after taking into account any potential class imbalances in the dataset because they are weighted by the number of true cases for each class.

4.4.2 Percentage Split

This study assesses how various training-to-testing splits affect a deep learning model's ability to recognize sounds related to home safety. A collection of 500 different audio recordings relevant to home safety were used in the dataset. For feature extraction, the model uses MobileNetV2, and for sound classification, it uses a bespoke classifier. In order to identify the best split for this application, three distinct data splits—70:30, 80:20, and 90:10—are looked at and their related model accuracies are analyzed.

Table 4.1 below displays how the dataset was divided into three distinct training-to-testing ratios to thoroughly evaluate the model's performance:

Table 4. 2 Percentage Split of Dataset.

#	Split	Percentage Ratio	Dataset	Accuracy
1	Training	70%	350	96%
	Test	30%	150	
2	Training	80%	400	93%
	Test	20%	100	
3	Training	90%	450	96%
	Test	10%	50	

The model was trained and then assessed on the corresponding testing sets for every split. The percentage of correctly identified occurrences in the testing set, relative to all instances in the testing set, was used as the primary evaluation parameter, or accuracy.

The findings show that the model retains a high degree of accuracy for various data divides, with 96% accuracy for the 70:30 and 90:10 divisions, respectively. It's interesting to note that the accuracy was somewhat lower (93%), with the 80:20 split.

With a 96% accuracy rate, the 70:30 split shows that a moderate size training set is enough to enable the model to learn from the data in an efficient manner, leading to strong performance on the testing set. This division offers a fair strategy, preserving a substantial testing set to precisely assess the model's generalization skills.

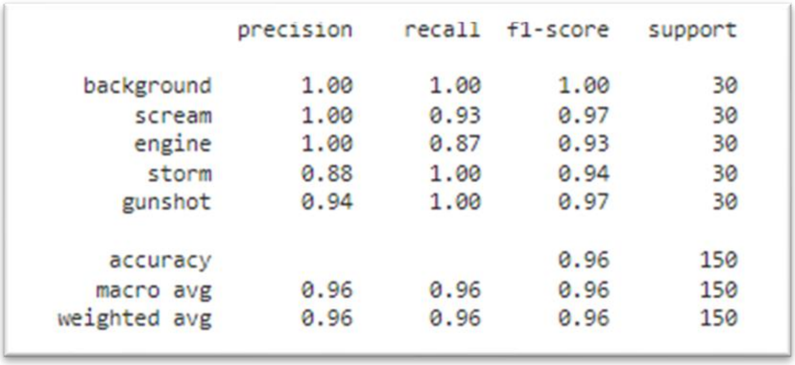
On the other hand, the lower accuracy of 93% for the 80:20 split raises the possibility that a smaller testing set could result in a less reliable performance rating. This finding emphasizes how crucial it is to have a sizable testing set in order to guarantee a thorough evaluation of the model's performance.

With a smaller testing set and a bigger training set, a 90:10 split can still provide excellent accuracy, as evidenced by its 96% accuracy rate. In contrast to the 70:30 split, the smaller testing set may not offer a more accurate assessment of the model's generalization performance.

The analysis of several data splits reveals that the 70:30 split is the most dependable and well-balanced option for training and assessing the deep learning model. It guarantees great accuracy while preserving a sizeable testing set for reliable performance evaluation. Though the smaller testing set of the 90:10 split results in a trade-off in the dependability of performance rating, the results are still promising. Even though the 80:20 split is still useful, its accuracy is a little bit lower, which emphasizes the importance of carefully taking testing set size into account when evaluating models.

4.4.3 Efficiency Based of Classification Report

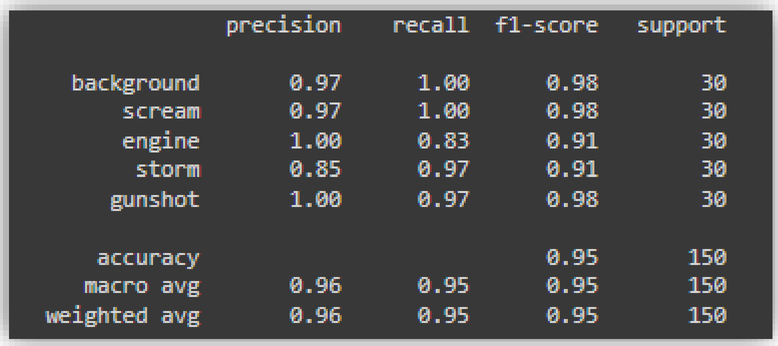
Three classification models that have been trained to distinguish five different sound classes—background, scream, engine, storm, and gunshot—are compared in this section. To assess the models' effectiveness, measures such as precision, recall, and F1-score were used. The data used for testing and training were divided 70:30. The performance metrics of each model are recorded and then a thorough analysis of their effectiveness follows.



```
              precision    recall  f1-score   support

  background       1.00      1.00      1.00        30
      scream       1.00      0.93      0.97        30
      engine       1.00      0.87      0.93        30
       storm       0.88      1.00      0.94        30
     gunshot       0.94      1.00      0.97        30

    accuracy                           0.96       150
   macro avg       0.96      0.96      0.96       150
weighted avg       0.96      0.96      0.96       150
```

Figure 4. 16 Model 1st Classification Report.

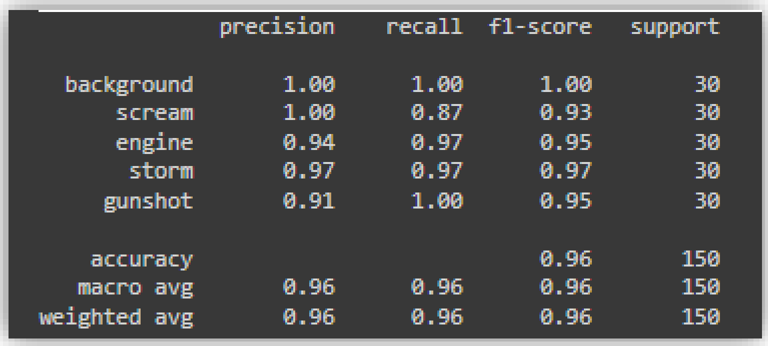
The results of the first classification model, shown in Figure 4.16 above, demonstrate good efficiency in all classes. For the background class, the model received an F1-score of 1.00, flawless precision, and recall. The scream class displayed an F1-score of 0.97, a recall of 0.93, and a precision of 1.00. The precision, recall, and F1-score for the engine class were 1.00, 0.87, and 0.93, respectively. The storm class's F1-score was 0.94, recall was 1.00, and precision was 0.88. In conclusion, the gunshot class received an F1-score of 0.97, a recall of 1.00, and a precision of 0.94. With weighted and macro averages of 0.96 for every parameter, Model 1 achieved an accuracy of 0.96 overall.



```
              precision    recall  f1-score   support

  background       0.97      1.00      0.98        30
      scream       0.97      1.00      0.98        30
      engine       1.00      0.83      0.91        30
       storm       0.85      0.97      0.91        30
     gunshot       1.00      0.97      0.98        30

    accuracy                           0.95       150
   macro avg       0.96      0.95      0.95       150
weighted avg       0.96      0.95      0.95       150
```

Figure 4. 17 2nd Classification Report.

Model 2nd provided constant performance with little modifications, as seen in Figure 4.17. It obtained an F1-score of 0.98, recall of 1.00, and precision of 0.97 for the background class. The F1-score for the scream class was 0.98, recall was 1.00, and precision was 0.97. The precision, recall, and F1-score for the engine class were 1.00, 0.83, and 0.91, respectively. The storm class's F1-score was 0.91, recall was 0.97, and precision was 0.85. With a precision of 1.00, recall of 0.97, and an F1-score of 0.98, the gunfire class performed well. With macro and weighted averages of 0.95, Model 2 achieved an overall accuracy of 0.95.



```
              precision    recall  f1-score   support

  background       1.00      1.00      1.00        30
      scream       1.00      0.87      0.93        30
      engine       0.94      0.97      0.95        30
       storm       0.97      0.97      0.97        30
     gunshot       0.91      1.00      0.95        30

    accuracy                           0.96       150
   macro avg       0.96      0.96      0.96       150
weighted avg       0.96      0.96      0.96       150
```

Figure 4. 18 3rd Classification Report.

The third model performed similarly to the first model with good efficiency in all classes, as shown in Figure 4.18. Perfect recall, F1-score of 1.00, and precision were all attained by the background class. The scream class displayed an F1-score of 0.93, a recall of 0.87, and a precision of 1.00. The engine class had an F1-score of 0.95, recall of 0.97, and precision of 0.94.

The storm class's F1-score was 0.97, along with precision and recall scores of 0.97. In conclusion, the gunshot class received an F1-score of 0.95, a recall of 1.00, and a precision of 0.91. With weighted and macro averages of 0.96 for every parameter, Model 3 achieved an accuracy of 0.96 overall.

After comparing the three models' levels of efficiency, models 1 and 3 are the most efficient, scoring flawlessly in the background class and performing well in the other classes—especially the scream and gunshot classes. While Model 3 displays balanced accuracy and recall for all classes, including better performance for the engine class as compared to Model 2, Model 1's slightly worse precision for the storm class is countered by flawless recall. Model 2 has a little lower recall for the engine class and a lower precision for the storm class, despite its steady performance. Overall, it appears that Models 1 and 3 are more efficient in sound classification tasks due to their high accuracy and balanced performance across all classes. In summary, all models exhibit good performance; however, Models 1 and 3 exhibit marginally higher efficiency. Even with its success, Model 2 still has room for development in several areas, suggesting that optimizing the training procedure or modifying the hyperparameters could increase the model's performance even more.

4.5 Conclusion

A sound classification system for home safety based on deep learning was effectively built and evaluated by the study. The entire process—from user input through a Streamlit interface to the output of classification results—was described in depth in the conceptual framework. Librosa and Matplotlib were used to convert audio files into spectrograms, which were then preprocessed and subjected to MobileNetV2 model analysis for feature extraction and classification.

Robust algorithm implementations and program codes allowed the program to reliably classify a variety of noises, demonstrating its efficacy. The CNN layers effectively processed these images to extract and classify features after the 'create_spectrogram' function effectively transformed audio recordings into Mel spectrograms.

With the ability to accept single, multiple, and real-time audio inputs, the prototype interface offered an intuitive user experience. The users could easily see the classification results, which included spectrograms and prediction probabilities.

After a rigorous evaluation, the model's accuracy was found to be high for all sound classes. Excellent recall, f1-scores, and precision were reported in a classification report; in particular, background noise and screams were distinguished with remarkable accuracy. The accuracy rate as a whole was 96%.

We looked at several training-to-testing splits to find the best data split for model evaluation. The most dependable split was found to be 70:30, which balanced testing and training data to guarantee accurate performance evaluation. The results showed that the 70:30 split is a reliable method for performance evaluation and that the deep learning model is very successful at categorizing sounds related to home safety.

Overall, the study showed how deep learning, web-based user interaction, and audio signal processing can be successfully combined to produce an accurate and effective sound classification system for home safety applications.

CHAPTER 5

CONCLUSION

5.1 Summary

In this project, convolutional neural networks (CNNs) are used to design and evaluate a sound detection system for home safety. In order to improve safety and security, the main goal of the study was to determine how well CNNs detect sound. Homeowners and building authorities were the target audience.

The significance of CNNs in enhancing the accuracy and efficacy of sound detection systems, notably in home security, was highlighted by a comprehensive literature study. In addition to sound detection, the review demonstrated CNNs' versatility and effectiveness, highlighting their promise for picture identification, natural language processing, and medical diagnosis, among other uses. This illustrated CNNs' wide range of applications and their fit for safe, intelligent living spaces.

The study used a systematic and organized methodology that started with a thorough literature review and problem statement. This initial round of preparation laid a solid basis for comprehending the difficulties and current approaches in sound event detection. The project had specific goals that included determining possible threats to home security, putting in place a CNN-based sound detection system, and assessing how well it worked.

500 audio samples from various sound categories related to machine and human activity were acquired as part of the data gathering process. During the design process, which included creating flowcharts, system architecture, and user interface designs, the goal was to maximize the CNN's capacity for sound event detection. A thorough implementation roadmap was given by the logistic regression pseudocode.

Software tools used in the prototype implementation were GoogleCollab for model training, VSCode for the user interface, draw.io for flowchart drawing, and Microsoft Word for documentation. Hardware-wise, an Acer Nitro 5 laptop with an Intel® CoreTM i5-8300 CPU, 8GB RAM, and Windows 10 was advised.

The focus of the system evaluation was on accuracy and efficiency. Accuracy measures included precision, recall, and F1 scores; efficiency measures included training and inference times, resource consumption, and real-time testing scenarios. The system's architecture, algorithms, training process, evaluation metrics, and optimization strategies were all covered in detail throughout the documentation phase. This information supported the iterative research methodology and recommended areas for additional improvement.

Through the use of an intuitive Streamlit interface, the deep learning-based sound classification system that was built showed excellent accuracy in classifying a variety of sounds. Librosa and Matplotlib were used to turn audio files into spectrograms, which were then examined by the MobileNetV2 model for feature extraction and categorization. The method successfully converted audio files into Mel spectrograms, which were then processed using CNN layers in order to extract and categorize features.

An easy-to-use interface was offered by the prototype, which accepted single, multiple, and real-time audio inputs. Viewers could see spectrograms and prediction probabilities along with the classification results. The model performed exceptionally well in differentiating between screams and background noise, with excellent accuracy across all sound classes. There was a 96% accuracy rate overall.

To find the best data split for evaluating the model, several training-to-testing splits were looked at; the most dependable split was the 70:30 split. This equal distribution of training and testing data ensures precise performance assessment. The outcomes validated how well the deep learning model classified sounds associated with home safety.

In summary, deep learning, web-based user interaction, and audio signal processing were effectively integrated in this research to produce an accurate and effective sound classification system for home safety applications. The research technique lays the foundation for future developments in the field of home safety sound detection by ensuring repeatability and transparency.

5.2 Contribution

Convolutional neural networks, or CNNs, were used in the creation and deployment of the Home Safety Sound Detection System, which has made numerous important contributions to society. These contributions include improving home security, developing technological applications for smart home systems, and laying the groundwork for further sound detection technology research and development.

1. Enhanced Home Security: This project's main contribution is to make homes more secure. Through efficient detection and categorization of diverse noises linked to possible security risks, such incursions, alerts, or emergency signals, the system facilitates prompt emergency responses. This helps to safeguard property, lessen the possibility that someone will be harmed, and create safer living conditions.
2. Homeowners and Building Authorities are Empowered: The system gives them a dependable tool for keeping an eye on possible safety hazards and taking appropriate action. People who feel more empowered may find greater peace of mind since they will know that their homes are being properly inspected for strange or harmful noises.
3. Better Quality of Life: person’s quality of life can be greatly enhanced by having the capacity to precisely identify and categorize sounds linked to home safety, especially for the elderly, the disabled, and single persons. The system can improve these people's sense of security in their homes by adding an extra layer of protection, which will improve their general wellbeing.
4. Educational Resource: Students and practitioners studying machine learning, audio signal processing, and smart home technologies can benefit from the project's thorough documentation and methodological approach. This project advances the knowledge and skills needed to advance these domains by offering a clear and comprehensive example of how to develop and assess a CNN-based sound detection system.

In conclusion, the Home Safety Sound Detection System created for this project makes significant contributions to society through raising the bar for smart home technologies, strengthening home security, laying the groundwork for more research, acting as a resource for education, and boosting people's quality of life. These contributions demonstrate how deep learning models can be effectively integrated into useful applications that tackle real-world problems.

5.3 Limitation

The Home Safety Sound Detection System has a lot of promise and efficacy, but throughout the research and testing stages, a few drawbacks were found. Resolving these issues is essential to improving the system's functionality and suitability for various real-world situations.

1. Limited Dataset Diversity: There were only 500 audio samples from categories related to human and machine behaviours that were utilized to train the Convolutional Neural Network (CNN). It's possible that this dataset doesn't include every sound found in different types of homes. The accuracy and generalization capacity of the system could therefore be impacted when it encounters unusual or uncommon sounds that aren't included in the training set.

2. Background Noise Interference: Even with the great accuracy rates attained, background noise interference could still affect the system. Ambient noise from things like traffic, the weather, or domestic appliances can make it more difficult to detect and classify objects in real-world contexts. This restriction may make it harder for the system to recognize important noises with accuracy.

3. Hardware and Computational Restrictions: An Acer Nitro 5 laptop equipped with an Intel® CoreTM i5-8300 CPU and 8GB RAM was used for the implementation and evaluation. Although this hardware setup is adequate for developing prototypes, performance constraints may arise when the system is implemented on devices with less processing power. Less powerful devices might find it difficult to meet real-time processing needs, which would limit the system's application in contexts with limited resources.

4. Problems with Scalability: The existing system is mostly targeted at individual residences and is tested and designed for a particular operational scale. The system may have difficulties with data management, processing power, and network infrastructure when it is expanded to bigger structures, apartment buildings, or community-wide safety

networks. The maintenance of accuracy and real-time responsiveness in the face of scalability would necessitate extensive adjustments and improvements.

5. Limitations of Real-Time Processing: The system's ability to process audio input in real-time is restricted to a certain amount of time, which might not be enough to record prolonged or sporadic sound events. This restriction may cause missed identifications or sluggish reactions in situations where noises happen frequently or over extended periods of time. For thorough monitoring, the system's capacity to manage continuous, lengthy audio streams would need to be improved.
6. User Interface and Experience: Although the prototype interface is meant to be easy to use, not all possible users may be able to fully utilize its features. It may be difficult for people with disabilities or low technical proficiency to interact with the system. For larger acceptance, the interface design must be improved to support a greater variety of user needs and preferences.
7. Variability in the Environment: The acoustics, layout, and occupancy of homes vary greatly. These elements may affect the accuracy of sound detection and propagation. For the system to properly adapt to various environmental conditions, it might need to be calibrated or customized. Maintaining robustness in a variety of household environments is still a major concern.

Even if the Home Safety Sound Detection System shows promise, these issues need to be resolved to maximize its effectiveness and guarantee its broad application. The dataset should be enlarged, background noise interference should be reduced, scalability should be improved, real-time processing should be improved, the user interface should be refined, and adaptability to different home situations should be guaranteed in future development.

5.4 Recommendations

Several recommendations are made to improve the Home Safety Sound Detection System's efficacy, scalability, and practicality in real-world scenarios, building on the knowledge and understanding obtained throughout its development and assessment.

1. Expand and Diversify the Dataset: The training dataset's growth and diversity should be given top priority in future endeavors. This entails adding a wider variety of sound types and situations found in various home settings. The system's capacity to generalize and precisely identify a broader range of sounds, including uncommon or unexpected events, will be enhanced by a larger dataset.

2. Using sophisticated noise reduction methods and adaptive algorithms can help lessen the negative effects of background noise on the accuracy of sound recognition. It is important to investigate techniques like dynamic noise modeling, adaptive thresholding, and signal filtering to improve the system's resilience in noisy settings without sacrificing real-time processing capabilities.

3. Optimize for Low-Resource Environments: The system must be optimized for low-resource environments in order to guarantee wider deployment feasibility. This entails lowering computing requirements, maximizing algorithm performance, and investigating neural network topologies that are lightweight and appropriate for use on edge devices with constrained memory and processing capacity.

4. Enhance Real-Time Processing Capabilities: It is imperative that the system be able to handle lengthy and continuous audio streams in real-time. To reduce latency and guarantee prompt detection and reaction to sound events, methods like parallel computing, stream processing, and effective buffering schemes should be researched.

5. Scalability and Network Integration: If the system is to be used for purposes other than private residences, it must be designed with scalability in mind. In order to do this, scalable infrastructures must be created, cloud-based data management and processing solutions must be integrated, and compatibility with current security and smart home networks must be guaranteed.

6. User-Centric Design and Accessibility: It is advised to improve the usability and accessibility of the user interface design in order to cater to a wide range of users. User experience and adoption can be improved by incorporating user feedback, performing usability testing, and putting intuitive features like voice commands or mobile app integration into the system.

7. Continuous Evaluation and Improvement: To sustain and improve system performance over time, a framework for continuous evaluation and improvement must be established.

This entails combining new developments in deep learning and audio signal processing research, assessing model performance continuously, and making frequent modifications based on user feedback.

8. Collaboration and Interdisciplinary Research: Interdisciplinary approaches will be promoted by fostering collaboration between researchers, practitioners, and stakeholders from a variety of domains, including machine learning, acoustics, psychology, and home security. Innovative solutions, cross-domain understandings, and comprehensive improvements in home safety sound detection systems can result from this partnership.

Future advancements can advance the state-of-the-art in-home safety technology and contribute to better and more secure living environments for individuals and communities by implementing these recommendations, building upon the foundation created by this research.

5.5 Conclusion

In summary, the utilization of Convolutional Neural Networks (CNNs) in the development and assessment of the house Safety Sound Detection System marks a noteworthy progression in augmenting house security via inventive technological implementations. The system's potential to create safer living environments is demonstrated by its ability to reliably identify and categorize a variety of noises linked to safety hazards. Nonetheless, there is room for improvement given several constraints, including dataset diversity, interference from background noise, and scaling problems. To increase the system's efficacy and applicability going forward, it is advised to increase the dataset, optimize for low-resource contexts, improve real-time processing capabilities, and improve user-centric design. The endeavours are designed to promote ongoing innovation and guarantee the dependability of the system in tackling changing obstacles related to home safety sound detection.

References

Abdoli, Sajjad, et al. “End-To-End Environmental Sound Classification Using a 1D Convolutional Neural Network.” Expert Systems with Applications, vol. 136, Dec. 2019, pp. 252–263, https://doi.org/10.1016/j.eswa.2019.06.040. Accessed 27 Mar. 2024.

Alzubaidi, Laith, et al. “Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions.” Journal of Big Data, vol. 8, no. 1, 31 Mar. 2021, journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8.

Banoula, Mayank. “Introduction to Long Short-Term Memory (LSTM) | Simplilearn.” Simplilearn.com, 2023, www.simplilearn.com/tutorials/artificial-intelligence-tutorial/lstm.

Bank, Drew. Image Segmentation Breakthrough: A New Era in CV - Labelify. 24 June 2023, www.datalabelify.com/en/image-segmentation-breakthrough-a-new-era-in-cv/.

Barroso, V., et al. “Applications of Machine Learning to Identify and Characterize the Sounds Produced by Fish.” Ices Journal of Marine Science, vol. 80, no. 7, 21 Aug. 2023, pp. 1854–1867, https://doi.org/10.1093/icesjms/fsad126.

Boesch, Gaudenz. “A Complete Guide to Image Classification in 2021.” Viso.ai, 24 Aug. 2021, viso.ai/computer-vision/image-classification/.

Carman, Ashley. “New IOS 14 Feature Lets the IPhone Alert You If It Hears Sounds like a Doorbell or Fire Alarm.” The Verge, 23 June 2020, www.theverge.com/21300261/ios-14-update-smoke-alarm-sound-detection-accessbility.

Daniel, B. K. “Older People’s Health Issues - Merck Manuals Consumer Version.” Merck Manuals Consumer Version, 2020, www.merckmanuals.com/home/older-people.

Davis, Kathryn L., and Donald D. Davis. “Home Safety Techniques.” PubMed, StatPearls Publishing, 17 July 2023, www.ncbi.nlm.nih.gov/books/NBK560539/.

DigitalJournal. “Voice Recognition Market: Global Trends and Prospects 2023.” Www.digitaljournal.com, 2023, www.digitaljournal.com/pr/news/prwirecenter/voice-recognition-market-global-trends-and-prospects-2023.

G. Priyadharshini, and Judie Dolly. Comparative Investigations on Tomato Leaf Disease Detection and Classification Using CNN, R-CNN, Fast R-CNN and Faster R-CNN. 17 Mar. 2023, https://doi.org/10.1109/icaccs57279.2023.10112860.

GeeksforGeeks. Understanding of LSTM Networks. 2020, www.geeksforgeeks.org/understanding-of-lstm-networks/.

Gysel, Philipp, et al. HARDWARE-ORIENTED APPROXIMATION of CONVO-

LUTIONAL NEURAL NETWORKS, https://arxiv.org/pdf/1604.03168.pdf.

Halil Ozgen Dindar, and Gökhan Dalkılıç. “Indoor Event Detection with Sound Data.” 2021 Innovations in Intelligent Systems and Applications Conference (ASYU), 6 Oct. 2021, https://doi.org/10.1109/asyu52992.2021.9599045.

Heittola, T., Virtanen, T., & Plumbley, M. Sound Event Detection: A Tutorial. 2021. https://arxiv.org/pdf/2107.05463.pdf.

IBM. “What Are Convolutional Neural Networks? | IBM.” Www.ibm.com, 2023, www.ibm.com/topics/convolutional-neural-networks.

Joshua, Paul, et al. Discovering the Optimal Setup for Speech Emotion Recognition Model Incorporating Different CNN Architectures. 1 Dec. 2022, https://doi.org/10.1109/hnicem57413.2022.10109279.

Kido, Shoji, et al. “Detection and Classification of Lung Abnormalities by Use of Convolutional Neural Network (CNN) and Regions with CNN Features (R-CNN).” IEEE Xplore, 1 Jan. 2018, ieeexplore.ieee.org/document/8369798.

Kousias, Konstantinos, et al. Long Short Term Memory Networks for Bandwidth Forecasting in Mobile Broadband Networks under Mobility, https://arxiv.org/pdf/2011.10563.pdf.

Lee, Mi-Young, et al. The Sparsity and Activation Analysis of Compressed CNN Networks in a HW CNN Accelerator Model. 6 Oct. 2019, https://doi.org/10.1109/isocc47750.2019.9027643.

Lee, Yejin, et al. “Prevention of Safety Accidents through Artificial Intelligence Monitoring of Infants in the Home Environment.” IEEE Xplore, 1 Oct. 2019, ieeexplore.ieee.org/document/8939675.

Mesaros, Annamaria, et al. Sound Event Detection: A Tutorial. 2021, https://arxiv.org/pdf/2107.05463.pdf.

Montaha, Sidratul, et al. “TimeDistributed-CNN-LSTM: A Hybrid Approach Combining CNN and LSTM to Classify Brain Tumor on 3D MRI Scans Performing Ablation Study.” IEEE Access, vol. 10, 2022, pp. 60039–60059, https://doi.org/10.1109/access.2022.3179577.

Morgan, Rachel, and Alexandra Thompson. Criminal Victimization, 2020. 2021. https://bjs.ojp.gov/sites/g/files/xyckuh236/files/media/document/cv20.pdf.

O. Hmidani, and M Ismaili. A Comprehensive Survey of the R-CNN Family for Object Detection. 12 Dec. 2022, https://doi.org/10.1109/commnet56067.2022.9993862.

Prachi Juyal, and Amit Kundaliya. Multilabel Image Classification Using the CNN and DC-

CNN Model on Pascal VOC 2012 Dataset. 14 June 2023, https://doi.org/10.1109/icscss57650.2023.10169541.

Purdy, Mark. “Sound Business: The Promise of Audio Machine Learning Technologies.” MIT Sloan Management Review, 28 Sept. 2023, sloanreview.mit.edu/article/listen-up-machine-learning-holds-great-promise-for-audio-applications/.

Qureshi, Rizwan, et al. A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023). 17 July 2023, https://doi.org/10.36227/techrxiv.23681679.v1.

Sugandhi, Abhresh. “A Guide to Long Short Term Memory (LSTM) Networks.” Www.knowledgehut.com, 7 Mar. 2023, www.knowledgehut.com/blog/web-development/long-short-term-memory.

Tholen, Celeste. “What Is a Home Safety Evaluation and How Do I Do One?” SafeWise, 27 May 2021, www.safewise.com/home-security-faq/home-safety-evaluation/.

Yanagisawa, Hideaki, et al. “A Study on Object Detection Method from Manga Images Using CNN.” 2018 International Workshop on Advanced Image Technology (IWAIT), Jan. 2018, https://doi.org/10.1109/iwait.2018.8369633.

Yun, Keon, et al. Behavior-Rule Specification-Based IDS for Safety-Related Embedded Devices in Smart Home. 1 Aug. 2021, https://doi.org/10.23919/wac50355.2021.9559588.

Zhang, Xue, et al. “A Brief Survey of Machine Learning and Deep Learning Techniques for E-Commerce Research.” Journal of Theoretical and Applied Electronic Commerce Research, vol. 18, no. 4, 1 Dec. 2023, pp. 2188–2216, www.mdpi.com/0718-1876/18/4/110, https://doi.org/10.3390/jtaer18040110.

Zhao, Jie. “Anomalous Sound Detection Based on Convolutional Neural Network and Mixed Features.” Journal of Physics: Conference Series, vol. 1621, no. 1, 1 Aug. 2020, p. 012025, https://doi.org/10.1088/1742-6596/1621/1/012025.