A NOVEL APPROACH FOR BREATHING SOUND DETECTION SYSTEM USING DEEP LEARNING

A Thesis Submitted
In Partial Fulfillment of the Requirements
for the Degree of

MASTER OF TECHNOLOGY

in

Artificial Intelligence

by

RAVI KANT SINGH
(Roll No. 2K22/AFI/30)

Under the Supervision of
Dr. RAHUL KATARYA
(Prof, Dept of Computer Science & Engineering)

To the
Department of Computer Science and Engineering

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042. India

May, 2024

ACKNOWLEDGEMENTS

I have taken efforts in this survey paper. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I am highly indebted to **Prof. Rahul Katarya** for his guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing this review paper. I would like to express my gratitude towards the **Head of the Department (Computer Science and Engineering, Delhi Technological University)** for their kind cooperation and encouragement which helped me in the completion of this research survey. I would like to express my special gratitude and thanks to all the Computer Science and Engineering staff for giving me such attention and time.

My thanks and appreciation also go to my colleague in writing the survey paper and the people who have willingly helped me out with their abilities.

Ravi Kant Singh

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

I, Ravi Kant Singh, Roll No. 2K22/AFI/30 student of M.Tech (Artificial Intelligence), hereby certify that the work which is being presented in the thesis entitled "**A Novel Approach for Breathing Sound Detection System using Deep Learning**" in partial fulfillment of the requirements for the award of the Degree of Master of Technology in Artificial Intelligence in the Department of Computer Science and Engineering, Delhi Technological University is an authentic record of my own work carried out during the period from August 2022 to May 2024 under the supervision of Dr. Rahul Katarya, Prof, Dept of Computer Science and Engineering. The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Place: Delhi **Candidate's Signature**

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor **Signature of External Examiner**

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE

Certified that **Ravi Kant Singh** (Roll No. 2K22/AFI/30) has carried out the research work presented in the thesis titled "**A Novel Approach for Breathing Sound Detection System using Deep Learning**", for the award of Degree of Master of Technology from Department of Computer Science and Engineering, Delhi Technological University, Delhi under my supervision. The thesis embodies result of original work and studies are carried out by the student himself and the contents of the thesis do not form the basis for the award of any other degree for the candidate or submit else from the any other University /Institution.

Prof. Rahul Katarya
(Supervisor)
Department of CSE
Delhi Technological University

Date:

A Novel Approach for Breathing Sound Detection System Using Deep Learning

Ravi Kant Singh

ABSTRACT

In this study a groundbreaking work on the detection of human breathing, a critical physiological function traditionally monitored by healthcare professionals is presented. The research introduces an innovative acoustic-based approach that leverages a deep learning model to discern the ultra-low intensity sound of respiration from ambient noise within a room. Utilizing a microphone, the model captures air pressure variations induced by respiratory activities, which are then transformed into a spectrogram image classification problem. The methodology involves applying a Fast Fourier Transform to the .wav file recordings, converting the time-series data into a frequency domain to visualize the subsonic breathing rates as spectrograms. These representations serve as inputs for a pre-trained ResNet18 model, enriched with additional layers through transfer learning, to identify breathing sounds with remarkable accuracy. The study displayed a significant 93% accuracy in detecting respiration within a 3-meter radius and has further developed into a multi-class classifier capable of estimating the distance from the sound source with an 87% accuracy. A bespoke dataset comprising 1700 instances with 13 classes has been curated for breath detection. The implications of this research are profound, offering substantial advancements in non-invasive monitoring techniques. The model's potential applications extend to disaster response and security monitoring, where it can significantly improve human detection in environments with poor visibility or restricted access. This thesis delineates the challenges, potential applications, and future research directions, marking a pivotal contribution to the interdisciplinary fields of biomedical engineering, electronics, and machine learning.

LIST OF PUBLICATIONS

1. R. K. Singh and R. Katarya, "Recent Trends in Human Breathing Detection Using Radar, WiFi and Acoustics," in *2023 6th International Conference on Recent Trends in Advance Computing (ICRTAC)*, IEEE, Dec. 2023, pp. 530–536. doi: 10.1109/ICRTAC59277.2023.10480776.

2. Paper title: 'Spectro-Acoustic Breath Detection – A ResNet18 Based Approach' is accepted in '2nd International Conferences on Optimization Techniques in Engineering and Technology (ICOTET 2024)' SCIE in May 2024.

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

Human breathing is a phenomenon that attracts the attention of various disciplines, such as science, medicine, and engineering. It serves as a crucial sign of health, a way to detect living people trapped under rubble, and a clue to possible diseases. Different fields use different methods to monitor and analyze human breathing for various purposes.

In the biomedical field, doctors use human breathing to check the health status of patients. They use devices like digital stethoscopes[1], which record the lung sounds using small microelectromechanical systems placed on the patient's chest. The recorded signals are then processed by signal processing techniques to estimate the respiration rate (RR) and other relevant parameters. Computer analysis and innovative technologies are used to examine the breath sounds and improve the diagnosis accuracy with machine learning methods. In the electronics field, radar waves[2] are used to detect human breathing in both line-of-sight (LOS) and non-line-of-sight (NLOS) conditions. This can be useful for disaster relief and security related applications. Post Covid-19 numerous researchers have used readily available Smartphone to capture patient's breathing sound and applied Deep Learning model to successfully discriminate positive samples.

However, in majority of the cases, smartphones were held in close proximity to the patient nostrils for better Signal to Noise Ratio (SNR). However, a microphone captures every sound generated in a room and thus, a machine learning model will be able to analyze recorded sound to detect human breath, thus discerning if a person is present in a room or not. A machine learning based model is hypothesized to overcome this challenge, that would detect the presence of hidden person in a room by analysing the surround sound consisting of ultra-low intensity sound of their respiration.

1.2 MOTIVATION

In areas prone to Insurgency and Anti-National Activities, security forces routinely under take area-search and house clearing drills. These operations are conducted by security forces as measure to counter insurgency, where they have to enter and secure buildings that may contain terrorists. Terrorists often hide in dark and covered areas of buildings, searching and locating such hideouts are a major challenge faced by security forces. To aid, security forces in dark, passive night vision-based devices are available. However, these devices are not effective in detecting persons hidden in LOS scenarios.

A possible solution may lie in exploiting the promising capability of Deep Learning[3] Models to estimate the possibility of a hidden human in dark hideouts, by processing the surround sound. Traditionally the speech part of sound is preferred over the background noise, which is ubiquitously discarded. However, these background sound noises can unlock possibilities that can help in detection of living beings trapped under debris of buildings, avalanches, etc. A specialized geometric arrangement of microphones can provide spatial information. Noise captured at such sites, upon decomposition into its sources and further localization can prove to be a life savior. With this motivation to uncover the components of noise, this survey of related technologies has been undertaken.

1.3 OBJECTIVES

(a) To carry out Literature Survey of breathing detection techniques.
(b) To prepare a smartphone/laptop recorded breath sound datasets.
(c) To develop a Deep Learning Model to detect human breath from smartphone/laptop recorded sound.

1.4 THESIS ORGANISATION

The thesis is structured into six distinct chapter as follows:

- Chapter 1 covers the motivation, introduces the topic, outlines the motivation behind the research question, and sets forth the objectives of the study.

- Chapter 2 offers a comprehensive review of the existing literature on methods for detecting respiration and their various forms. It also highlights the limitations of these methods and examines the datasets employed in prior research.
- Chapter 3 makes the case for the creation of a new dataset and details the assumptions, environmental conditions, and strategy for the collection of respiratory data.
- Chapter 4 is dedicated to the development of the model, including data preprocessing and the implementation process.
- Chapter 5 evaluates the results achieved and discusses the hyperparameters used in the experimental framework.
- Chapter 6 wraps up the study, discussing potential future research and the societal and industrial relevance of this specialized field.

CHAPTER 2

RELATED WORK

2.1 SYSTEMATIC SURVEY ON BREATHING DETECTION

Detecting human respiration is crucial for several sectors, including health care, security, and emergency response. Each sector has unique demands for detecting breath, such as precision, dependability, affordability, ease of use, scalability, and practicality. Additionally, the detection process can be affected by a range of elements like ambient noise, the subject’s movements, interference from cardiac sounds, or compatibility with devices. As a result, a variety of approaches have been devised, falling into two primary groups: contact and non-contact methods as shown in fig 2.1. Contact methods require the subject to be in direct touch with an instrument, like a stethoscope, microphone, or sensor. Non-contact methods, on the other hand, do not require touch and instead employ wireless technologies like radar, WiFi, or ultrasound to detect and gather respiratory data.

2.1.1 Essential Terms

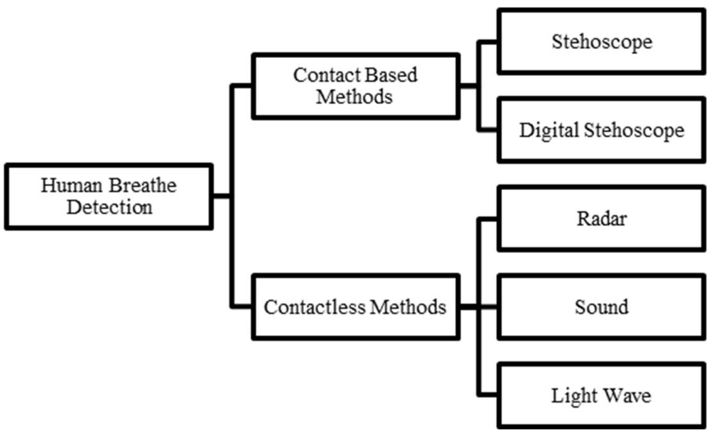

Figure 2.1: Human Breath Detection Techniques

In this section we briefly covered few essentials’ terms which will be required in subsequent sections.

(a) Exhalation - Exhalation is the release of air from lungs. This sudden egress of air creates a very faint perturbation in surrounding air. A sophisticated equipment like a stethoscope or diaphragm of a microphone is capable of observing the same. Microphones, on the other hand, are used for recording sounds. Faint sound like breathings is recorded. Human breathing sounds (including nasal and lung breathings) have provided essential insights into persons health. Markandeya et al [4] have demonstrated that smartphone speakers are quite effective at close range in detection of breathing patterns and subsequently identification of potential diseases.

(b) Reference Sound Pressure - A perturbation in air travels at the speed of sound in the medium. Perturbation results in a small change in air density at a given point in space at any time. The smallest perturbation in sound pressure level measure by equation (1) that aligns with the threshold of human hearing is 20 micro pascals[5]. It is also the reference value sound pressure p_0. The sound pressure level for p Pa in dB is given by equation

$$dB = 20log\frac{p}{p_0} \tag{1}$$

(c) Human Hearing Threshold - The minimum value of sound pressure, at which a sound of the given frequency can be perceived by a human ear – hearing threshold – is an individual property depending on listener and his age.

(d) Breath Sound - It includes normal and adventitious sounds[6], [7] recorded over the chest wall, the trachea or at the mouth. Their generation is related to airflow in the respiratory tract. Acoustically, this sound is characterised by a broad-spectrum noise with a frequency range depending on the pick-up location.

(e) Types of Microphones

i) **Condenser Pressure Microphone** - These microphones have a movable diaphragm and a fixed plate, together these function as capacitors[8]. The capacitance changes as the diaphragm moves back and forth. The diaphragm responds to variations in air pressure.

ii) **Magnetic Induction Microphone** - In this type, a small metallic film is suspended in an electromagnetic field[9]. The metal film moves in response to variation in air pressure, inducing a current in the circuit that is proportional to air pressure change.

iii) **Electret Microphone** - Smartphones typically use electret capacitor pressure microphones. Electret is a pre-polarized material and capacitor backplates are typically coated with it, resulting in excellent performance at low cost. which are condenser pressure microphones, built over an integrated circuit.

iv) **Ambisonics microphones** - These are omnidirectional microphones, which capture sound from all directions[10]. It uses 4-channel, and also stores the spatial information of sound.

2.1.2 Breathing Detection Techniques

(a) Traditional Techniques

i) **Doctor's Ear.** Historically, the practice of clinical diagnosis has relied on a physician's verification of particular respiratory sounds and the examination of thoracic organs through auscultation. A physician would listen directly to a patient's body to assess a targeted organ.

ii) **Stethoscope.** The original stethoscope was a basic device, comprising a hollow wooden tube and a membrane that spanned its entire working surface. This design served to amplify the sound and increase its volume. The air within the device's acoustic channel functioned as a medium for sound conduction, meaning that any obstruction or compression within this flexible channel could disrupt the auscultation process. Nonetheless, the technique of lung auscultation is inherently subjective and prone to significant variation. Such variability often stems from the examining physician's level of experience and expertise, as well as various patient-related factors. Presently, the medical fields of acoustics, informatics, and artificial intelligence are grappling with the challenge of making the analysis of respiratory sounds more objective.

iii) **Digital Stethoscope.** A digital stethoscope transforms acoustic sounds into digital data that can be displayed, enhanced, captured, archived, and analyzed using computer systems. These stethoscopes capture respiratory sounds by positioning a compact microelectromechanical system (MEMS) recording device on the patient's chest. The digital recordings are then subjected to signal processing methods to calculate the breathing rate (RR).

(b) Sound Based Techniques

One of the contactless methods involves capturing breathing sound at some distance from the person and then processing it for further investigation. This method has the advantages of providing additional information besides respiration rate, thus is widely researched in medical informatics field. In these methods respiratory sounds at the nose, mouth, or trachea is recorded that has been used for various purposes by applying different ML and DL methods as covered in Table I.

i) **Polysomnogram.** Polysomnography is a type of sleep study that measures various physiological parameters during sleep, such as brain activity, eye movements, muscle activity, heart rhythm, breathing, and oxygen levels. The test result is called a polysomnogram, which shows the sleep stages and cycles, and can help diagnose or rule out many sleep disorders. A polysomnogram is a file that contains the data from a polysomnography, which is a sleep study that measures various physiological parameters during sleep. The file type of a polysomnogram depends on the software and device used to record and analyze the sleep data. Some common file types are:

- **EDF**: European Data Format, a standard file format for biomedical signals that can store multiple channels of data with different sampling rates and resolutions.
- **PSG**: Profusion Sleep Data Format, a proprietary file format used by Compumedics, a company that produces sleep diagnostic systems. PSG files can store up to 256 channels of data, as well as annotations, events, and scoring information.

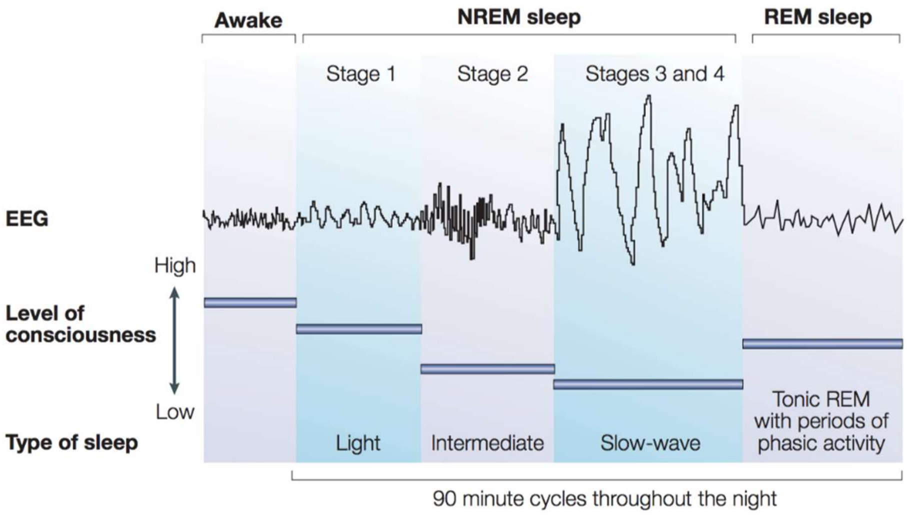

Figure 2.2: A Polysomnogram Recording[37].

A polysomnogram file as shown in Fig 2.2 can provide valuable information for diagnosing and treating various sleep disorders, such as obstructive sleep apnoea, narcolepsy, restless legs syndrome, and insomnia.

ii) **Feature Extraction Techniques.** Feature extraction techniques for audio signals are methods to extract useful information from audio data, such as speech, music, or sound effects. Feature extraction is used for various applications, such as audio classification, speech recognition, music analysis, audio segmentation, and more. Some common feature extraction techniques for audio signals are:

- **Temporal features**: These are features that are computed from the audio waveform itself, without any transformation to the frequency domain. Examples of temporal features are zero-crossing rate, energy, root mean square, autocorrelation, etc.

- **Spectral features**: These are features that are computed from the frequency representation of the audio signal, such as the Fourier transform, the short-time Fourier transform, or the mel-frequency cepstrum. Examples of spectral features are spectral centroid, spectral rolloff, spectral flux, spectral flatness, spectral entropy, etc.

- **Cepstral features**: These are features that are computed from the cepstrum of the audio signal, which is the inverse Fourier transform of the logarithm of the spectrum. Examples of cepstral features are linear predictive coding, mel-frequency cepstral coefficients, perceptual linear prediction, etc.

- **Harmonic features**: These are features that are computed from the harmonic structure of the audio signal, which is the set of frequencies that are multiples of the fundamental frequency. Examples of harmonic features are harmonic ratio, inharmonicity, pitch, etc.

- **Rhythmic features**: These are features that are computed from the rhythmic pattern of the audio signal, which is the temporal arrangement of beats, onsets, or events. Examples of rhythmic features are tempo, beat histogram, onset detection, etc.

iii) **Auto Regressive.** An autoregressive model is a statistical model that predicts future values of a variable based on its past values. It is used to analyze and forecast time series data, such as stock prices, economic indicators, weather patterns. Since audio is a time series data, temporal features extracted from acoustics data are frequently used with Auto Regressive models.

iv) **Welch Periodogram.** A periodogram is an estimate of the spectral density of a signal, which shows how the signal's power is distributed over different frequencies. It is calculated by obtaining the Fourier transform of the signal and then squaring it off. A periodogram analyzes the periodic or cyclical behaviour of a time series. Welch Periodogram involves dividing the original signal into overlapping segments, windowing each segment, computing the Fourier transform of each segment, and averaging the squared magnitudes of the results. This reduces the overall noise and variance of the estimate compared to the traditional method.

v) **Fast Fourier Transform.** Fast Fourier Transform (FFT)[11] is a mathematical algorithm that computes the Discrete Fourier Transform (DFT) of a signal. The DFT transforms a signal from the time domain to the frequency domain, thus, it shows how the signal's power is distributed over its constituent frequencies. FFT in audio signal processing acts as a feature extractor to analyze and manipulate the frequency components of the audio signal. For example, FFT can be used to:

- Visualize the musical notes, chords, and pitch of the audio signal by plotting the spectrum or spectrogram of the signal.
- Filter out background noises, unwanted frequencies, or harmonics by applying a low-pass, high-pass, band-pass, or notch filter to the signal in the frequency domain.

Table. 2.1. Different purposes for Respiration Rate Detection

Purpose	**Model**	**Datasets**	**Remarks**
Detecting COVID-19 using respiratory sounds	Alkhodari M et al (2020) [38]demonstrated that deep learning along with hand-crafted features achieved	480 samples from COSWARA	94.5% discrimination results for COVID-19.
	Campana M et al (2022) [39]demonstrated a L3 Net with transfer learning to automatically extract meaningful features	COSWARA, Virufy, Cambridge	Superior performance
Classifying respiratory sounds into different types or diseases	Yu S et al (2022) [40]Glance and Glaze neural network	ICBHI 2017 database	84.7% accuracy
	Kranthi et al (2022)[41] regularized deep convolutional neural network	KDD-data, ComParE2021-CCS-CSS-Data, and NeurlPs2021-data	90-94% accuracy
	Wang L et al (2023) [27]applied an agglomerative hierarchical clustering algorithm	212 PSG Audio, contains tracheal sound	80-88 % accuracy
	Furman G et al (2022) [42]applied fast Fourier transform (FFT) over respiratory sound recordings for diseases	56 patients made using a smartphone	86 % accuracy
Estimating respiration rate or sleep stages using respiratory sounds	Yunyoung et al (2016)[26] estimated the respiration rate from nasal breathing recordings	10 Subjects	1% of mean absolute error (MAE).
	Fahed V et al (2023) [43]estimated the respiration rate from audio recordings made by a smartphone.	112 Participants sound	MAE of 0.63 bpm
	Dohney E et al (2023) [44]estimated the respiration rate from audio recordings made by a smartphone.	220 Patient recording	MAE of 0.2 bpm
	Hong J et al (2022) [25]used an LG G3 smartphone to record nocturnal sounds	327 patients from a distance of 1m	sleep stages with 68% accuracy
	Tran H et al (2023)[45] trained a deep learning architecture to predict sleep stage	829 Smartphone recording at home	76.2 % accuracy

(c). Radar Based Techniques

It is one of the common methods to measure human breathing is to use radar waves to either directly or indirectly probe the target and observe the doppler shift in the reflected waves. Bugaev E et al [2]proposed one of the earliest methods to detect human heart rate using short-range radar waves in 2006. They suggested a rejection method for the probing and reflected signals, which had high sensitivity to detect objects with vibrating boundaries. Many researchers have used the doppler effect to detect the chest vibrations of humans. Li W [15]et al successfully detected breathing up to 1m using micro-doppler based on Cross Ambiguity Functions in 2016. Dellversana et al[16] used 10 GHz

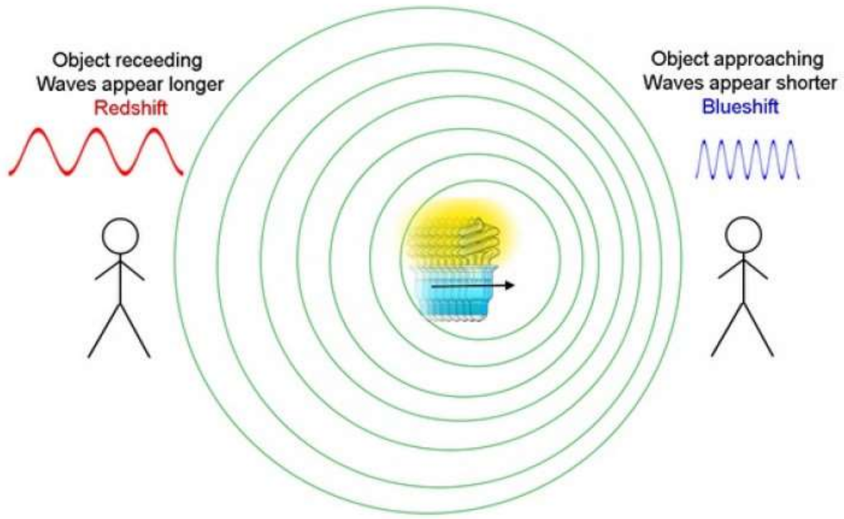

Figure 2.3: A diagrammatic representation of Doppler Effect[46].

mmWave Continuous Wave to detect human breathing through walls. However, using mmWave radar for probing can be expensive.

i) **Doppler Effect.** The Doppler effect[17] in radar is the phenomenon that causes a change in the frequency of the radar signal when the target is moving relative to the radar as shown in Fig 2.3. The Doppler effect can be used to measure the velocity and direction of the target, as well as to filter out stationary clutter and noise. The Doppler effect depends on the speed of the target, the frequency of the radar signal, and the angle between the radar and the target.

ii) **Frequency Modulated Continuous Waves.** Cardillo E et al [18] simulated a solution using two different mmWave radars operating at 24 GHz and 122 GHz. Li G et al [19] proposed a FMCW radar operating at 77 GHz, which was effective for NLOS situations and also enabled pose estimations. Ultra-Wide band has been effectively used for monitoring Sleep Breathing Pattern by Husiani M et al[20].

iii) **WiFi.** Therefore, some researchers have explored low-cost solutions based on WiFi Antenna and Software Defined Radios (SDRs)[15]. Costanzo et al[15] implemented a solution based on SDRs, where the subject was placed a few centimetres from the antenna. Few researchers demonstrated that WiFi antennas placed in buildings, can also be effectively employed to monitor respiration rate. Chen C [21] et al used the Channel State Information (CSI) of WiFi Antennas to capture the small variations in the environment caused by breathing. Guan L et al [22] extended this idea by using Beam Forming technology to scan a room and detect abnormal breathing in multi-person situations using WiFi Antenna. Their method achieved 98.07% accuracy in detecting nine people in a Non-Line of Sight (NLOS) scenario.

iv) **WiFi Channel State Information.** WiFi CSI is a measure of the properties of the wireless channel between a transmitter and a receiver. It describes how the WiFi signal propagates and reflects in the environment, and how it is affected by factors such as distance, interference, fading, and noise. CSI can be used for various applications, such as WiFi sensing, localization, beamforming, and rate adaptation.

(d). Light Based Techniques

i) **Smartphone Camera.** Processing video clips and scanning finger using front camera of smartphones has also helped researchers estimate respiration rate and heart rate respectively. Bae S et al [23] have proposed a novel algorithm of estimating respiration rate by using a video of a person. Chest movements of the persons are measured to estimate his respiration rate with a MAE of 0.78 bpm.

In the Table 2.2, a brief description of work that has been done in the past is given. A special emphasis has been given to cover techniques that have used smartphones to record breathing sound.

Table 2.2- Breathing Detection Techniques studied in this survey

Ref	Wave	Research Problem	Advantages	Key Features
[22]	EM Wave	Multi Person Breathing Detection	Multi Person Breathing, Abnormal Breathing.	WiFi Antenna
[43]	Sound Wave	Random Forest based Classification of Breathing Phase from Audio recording	MAE of 0.63 bpm	Smartphone
[47]	Sound Wave	Contactless Breathing Detection Using Smartphone	MEE 0.17 bpm	Xiaomi NIX2, Google Nexus 5
[44]	Sound Wave	Estimation of RR using audio signals recorded by Smartphone	MAE of 0.2 ± 0.27 bpm	Smartphone
[45]	Sound Wave	Predicting sleep stages using Smartphone audio recording in Home Environment	accuracy of the model was 76.2%	Smartphone, Home Environment
[27]	Sound Wave	Tidal Volume Estimation using Respiratory Sound	80-88 % accuracy	Tracheal Sound
[19]	EM Wave	Detection of Human Breathing in Non-LOS using mmWave FMCW	Avg Error 1% in RR estimate, Pose Estimation	77GHz
[20]	EM Wave	Sleep Breathing Detection	Lowest avg error of 6.12	UWB Radar
[24]	Sound Wave	Timely detection of abnormal respiratory sounds such deep and heavy breaths	Tiny model on Edge devices: 91 % f1 score	Samsung Galaxy Note 8
[39]	Sound Wave	Covid 19 Detection from Smartphone Audio Data	superior performance	Transfer Learning, L3 Net
[38]	Sound Wave	Detection of Covid 19 in smartphone-based breathing recordings	COVID 19 discriminatory accuracy 94.5%	Smartphone recording
[25]	Sound Wave	End-to-End sleep staging using nocturnal sounds from microphone chips	68% accuracy	LG G3, 1m from participant
[41]	Sound Wave	Respiratory Disease diagnosis with RDCNN using respiratory sounds	90-94 % accuracy	work with cough, voice sound
[42]	Sound Wave	Remote analysis of respiratory sounds in Patients with Covid19	Process on High Freq part	Smartphone, near mouth
[18]	EM Wave	Radar range, breathing separation and detection of human in a cluttered environment	Multiperson, 3d Map	24 GHz, 122GHz
[48]	Sound Wave	Recognising changes in breathing pattern for disease detection.	0.5 sec time for processing whole breathing sequence.	Microphone near nose 2cm
[15]	EM Wave	Software Defined Radio for Breathing Sound Detection	Low cost, easily optimized	Upto few centimetres
[4]	Sound Wave	Can smartphone-based microphone be used for detection and classification of upper airway obstructions.	Smartphone microphone in range (20Hz -22kHz)	Samsung Galaxy S3,
[21]	EM Wave	Time Reversal: Breathing rate detection and estimation	accuracy of 98.07% in breathing rate estimation	WiFi Antenna, nine people under the NLOS
[16]	EM Wave	Through the wall breathing detection by means of doppler radar	clear and peaked frequency estimations	mmWave CW, 10GHz, Upto 1m
[49]	EM Wave	Non-Contact Breathing Detection	LOS, NLOS, LF Narrow Bandwidth	Upto 1m from Participant
[26]	Sound Wave	Estimate Nasal Breathing Rate using sound recording from a smartphone	Median error of 1%. Upto 30 cm.	iPhone 4S
[32]	Sound Wave	Breathing detection using acoustic sensor	91.3 % accuracy	Microphone on Neck

2.1.3 Dataset Studied

In the Table 2.3 we have discussed different datasets that have been used in papers covered in Table 2.2 for human breathing detection and analysis. None of the datasets covered below are specifically designed to detect human breathing in a room.

Table 2.3 – Datasets covered in this Literature Survey

Author	Dataset
Mrunal N Markandeya[4]	700 patients snoring sound,
Rong Phoophuangpairoj[48]	400 wav files
Tran-Anh D[24]	BreathSet (1157 PSG),
Campana M[39]	Coswara[7], Vurify
Alkhodari M[38]	Coswara,
Hong J[25]	1157 PSGs, Smartphone 327 patient
Kranthi Kumar[41]	KDD, ComParE2021-CCS-CSS, and NeurlPs2021
Furman G[42]	56 Patients
Yu S[40]	ICBHI 2017
Fahed V[43]	112 Participants Respiratory signals
Doheny E[44]	220 Patient Recording
Tran H[45]	829 Smartphone
Wang L[27]	212 Polysomnograms
Bae S[23]	50 Patient in RR

These datasets vary in size, format, source, and application. A summary of the comparison of these datasets is:

(a) **Size**: The datasets vary in size, from 27 COPD patients in BreathSet [24]to 1157 PSGs from 327 patients in Hong J et al[25]'s dataset. The largest dataset is Coswara, which contains raw respiratory audio of more than 20,000 subjects.

(b) **Source**: The datasets have different sources, such as smartphones, microphones, or stethoscopes. Most of the datasets use smartphones to record the respiratory sounds, such as Yunyoung et al [26]& Mrunal et al[4]. Few datasets use microphones to record the respiratory sounds, such as Wang L et al[27], and Semmad A et al[28].

(c) **Format**: The datasets have different formats, such as wav files, audio samples, or polysomnograms. Some of the datasets use audio samples to store the respiratory sounds. Some of the datasets use polysomnograms to store the respiratory sounds, such as Hong J et al[25] and Bae S et al[23].

(d) **Application**: The datasets have different applications, such as detecting COVID-19[29], [30], classifying respiratory sounds into different types or diseases, estimating respiration rate or sleep stages, identifying living persons under debris or hidden security threats behind walls, and monitoring health and disease stages.

2.2 LIMITATIONS IN EXISTING WORKS

2.2.1 Findings

(a) Active methods. These methods use high energy EM waves in LOS and NLOS scenarios for breathing detections. However, they require specialized hardware and have high power consumption[16]. Alternatively, WiFi based methods have low power consumption, but are inherently dependent on a reliable and effective WiFi infrastructure[22]. Latter is not guaranteed to be active in disaster or present in a non-urban area.

(b) Passive methods. Capturing breathing sound are least intrusive, least power dependent and most economical for wide-spread use. But it poses its own key challenges:

i) Sound waves pressure level reduces as it travels away from the sources[5].

ii) Breathing Sounds have very low frequency, and low intensity[31].

iii) They are more prone to multitude of disturbances such as from ambient noise or temperature or subject’s movement or posture.

iv) In every research conducted so far, breathing sound has been captured by placing the microphone in close proximity to the subject[32], [33].

v) None of the research paper have targeted to detect breathing, but have focussed on consequent task of respiration rate estimation, sleep stage monitoring.

(c) Datasets. None of the datasets given in Table III has been captured by placing microphone farther than 30cm from the source. The datasets presented in Table III pose some challenges as many are curated for disease detection or consist of polysomnogram recordings. Some limitations are:

i) **Generalisability**: The datasets may not represent the diverse and heterogeneous population of human subjects, as they have different age, gender, ethnicity, health status, or environmental conditions.
ii) **Consistency and compatibility**: The datasets may not be consistent or compatible with each other or with the devices and software for recording and processing. For example, some datasets have different formats, sources, applications.
iii) **Accessibility**: Some datasets may require special permissions, licenses, or agreements to access them. Some datasets may not be publicly shared or published due to privacy or security reasons.

2.2.2 Challenges

This study identifies following main challenges in detecting human breath from its nature, surroundings, and recording device:

a) **Subject-related**: Respiratory sound recording can be difficult due to mask wearing, subject movement or posture, distance from microphone or sensor, or lack of cooperation. These can affect signal-to-noise ratio, signal quality, or signal availability, which can impact the methodology's feasibility and performance.

b) **Surrounding-related**: Respiratory sound recording can also be influenced by sensor position, background noise, or other surround sounds in the environment. These can introduce noise, distortion, or ambiguity in the signals, which can reduce the analysis' accuracy and reliability.

c) **Recording device-related**: Respiratory sound recording can be limited by the inherent characteristics of the recording device, microphone or sensor quality, its audio processing techniques.

2.3 PROBLEM STATEMENT

Apropos the ability of microphones fitted in smartphone to discern human breathing sound has been clearly elucidated by various researchers and their work in Table II. However, in this work, our key observations are:

a) Microphones were placed in close proximity to source (between 1cm to 30cm)
b) Audible breathing, coughing, sneezing etc sounds were targeted by the researchers.
c) Their research highlighted that aspect of microphones ability to discern breathing sound upto 30cm only.

Contrary to the aforementioned research, our goals encompass not only audible but also inaudible respiratory sounds, with the intention to identify breathing from a distance exceeding 30cm. Therefore, we have refined our objective into the following problem statement:

a) To compile a dataset comprising inaudible respiratory sounds detectable up to a distance of 5 meters.
b) Utilizing a supervised learning approach, develop a deep learning algorithm capable of identifying inaudible respiratory sounds, based on a labeled dataset.

2.4 APPROACH TO SOLUTION

Respiratory activities induce variations in ambient air pressure, manifesting as sound waves with alternating compression and rarefaction zones. The process of respiration engages the vocal cords, trachea, and mouth as vibrational sources, instigating the oscillation of air molecules along the path of the wave. These oscillations facilitate the transmission of sound waves without causing a net displacement of the molecules.

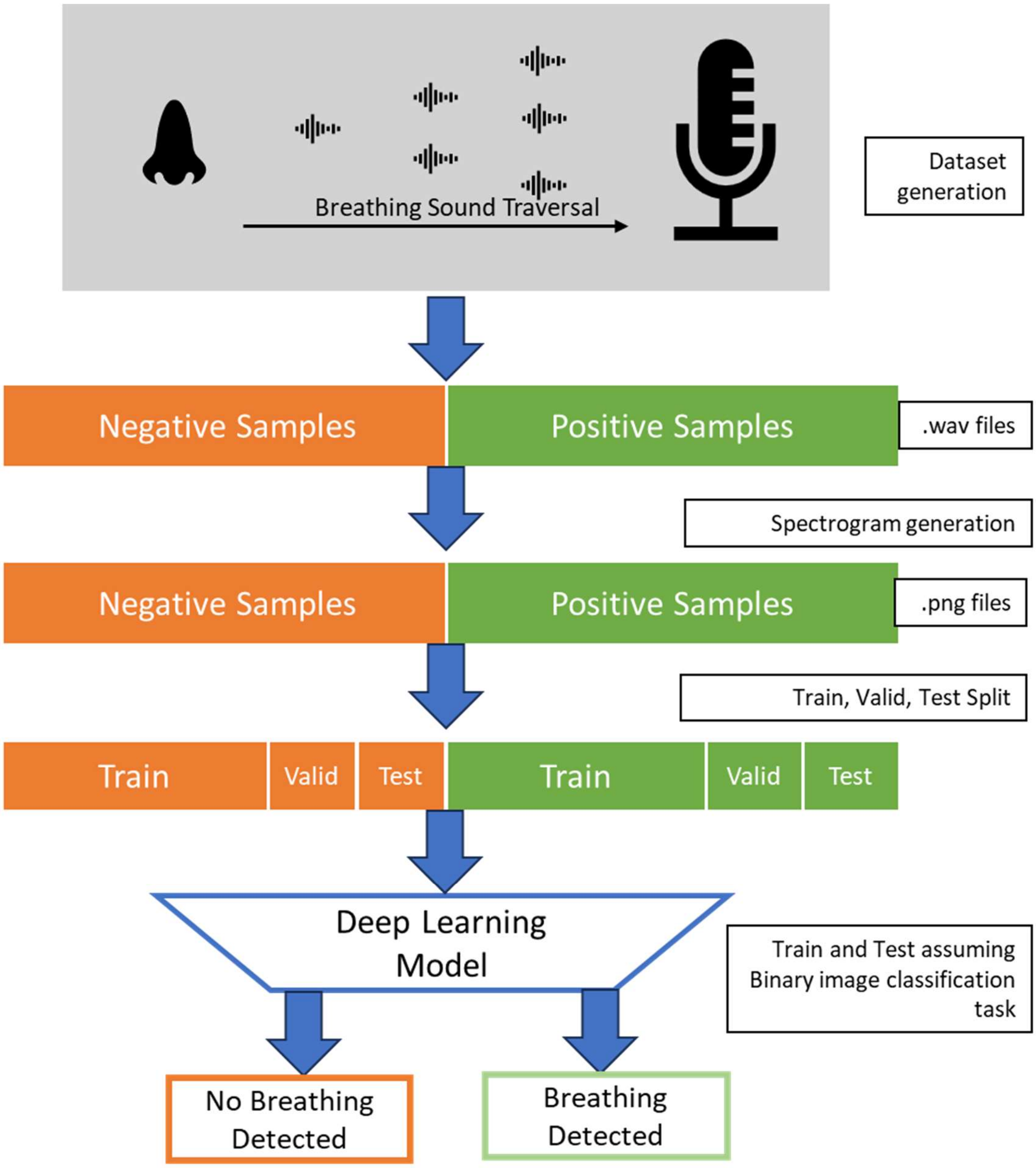

Figure 2.4: Flowchart of Proposed Approach to Breathing Detection Problem

The microphone captures these pressure variations over time, resulting in a '.wav' file as shown in Fig 4 that exhibits the temporal fluctuations of pressure at the microphone's diaphragm. Applying a Fast Fourier Transform to this time-series data reveals its constituent frequencies.

Humans typically breathe at a rate of 5-20 breathing cycles per minute, corresponding to a frequency below 1Hz, which is considered subsonic. A spectrogram visualizes these constituent frequencies and their respective amplitudes over time as a heatmap, effectively serving as a three-dimensional array analogous to images.

These 'images' can be labeled to indicate the presence or absence of breathing sounds and then input into pre-trained deep learning models to detect the presence of breathing sounds within them, as illustrated in the flowchart in fig 4.

Employing a strategic microphone arrangement can yield spatial data, enabling the detection of living entities through sound analysis. Our research concentrates on the acoustic signature—specifically, the breathing of living beings—to address the challenge of detection in environments with low visibility or that are otherwise inaccessible. This capability is vital for search and rescue missions and for surveillance of restricted areas. Through the examination of the spectral characteristics of respiratory sounds, we have devised a binary classification system that proficiently identifies human presence via their breathing sounds.

CHAPTER 3

DATASET-BRTHSOUNDSET

In this subsection, we present the rationale behind the creation of a new dataset, describe the methodology of its compilation, and discuss the foundational principles guiding this initiative.

The primary goal was to create a new dataset that enables the detection of human respiration within inaudible acoustic spectra. Existing datasets, as listed in Table 2.3 and reviewed in our previous research[34], are characterized by audible respiration sounds, which aren't suited with this paper objectives.

These pre-existing datasets were primarily created for diagnosing respiratory anomalies or sleep disorders, and they were not tailored for the specific task of detecting human respiration at a distance. They include recordings of respiratory sounds taken at close range, from 1 cm up to 30 cm from the nostrils, which is a significant limitation for the current applications.

To address this gap, a unique dataset, '**BrthSoundSet**' was created. It is specifically designed for the detection of human breath within acoustic signals.

3.1 ASSUMPTIONS

In the process of creating a dataset for the detection of human respiration via acoustic waves, we posited several assumptions to streamline the dataset and ensure its pertinence and practicality. These presumptions include:

a) **Breathing Sound Characteristics**: We postulate that respiratory sounds possess distinct acoustic signatures that can be discerned from other ambient sounds.
b) **Microphone Sensitivity**: The microphone is presumed to possess sufficient sensitivity to accurately document the subtleties of human respiratory sounds across varying distances.

c) **Environmental Consistency:** It is assumed that the recording milieu maintains a steady ambient noise level that does not substantially impede the respiratory sounds

d) **Breathing Patterns**: We anticipate that the recorded breathing patterns will exhibit a spectrum of respiratory rates and depths, thereby simulating an array of breathing scenarios

e) **Data Integrity:** We assume the data procured is devoid of corruption and faithfully represents the respiratory sounds without significant quality degradation

f) **Labelling**: Two labelling paradigms have been instituted; one for binary classification, denoting presence (1) and absence (0), and another for multi-class classification, where label '0' implies no person and any other numerical value denotes distance between source and recording device in 'cm'.

3.2 DATASET ACQUISITION STRATEGY

Our dataset collection strategy was meticulously designed to ensure a balanced composition of positive and negative samples.

3.2.1 Negative samples are those without any human presence. Due to the practical challenges of confirming absolute human absence during ambient sound recording, we classified a sample as 'Negative' if no person was detected within a 5-meter radius of the recording device. We successfully gathered 500 such negative samples.

3.2.2 Positive samples consist of sets where a single individual's presence is recorded, with their distance from the recording device encoded in the filename, as depicted in figure 5.

To infuse diversity into our dataset, we captured respiratory sounds under two distinct conditions: Line of Sight (LoS) and Non-Line of Sight (NLoS).

(a) LoS: We recorded 500 audio samples with a subject positioned directly in front of the recording device at intervals of 30cm, 60cm, 100cm, 200cm, and 300cm, ensuring 100 samples for each specified distance.

(b) NLoS: This category was expanded by differentiating scenarios where the subject's orientation was away from the recording device while remaining in the same room. Further diversification was achieved by placing the subject outside the room with the door either partially or fully closed, in accordance with the data acquisition methodology outlined in Fig 3.1. The chart specifies that all cells under "No of Samples" contain the number '100', indicating the uniformity of sample size across different conditions and distances.

Figure 3.1: Dataset Acquisition Plan for BrthSoundSet

For the creation of this dataset, I personally undertook the role of the human subject and meticulously labelled each dataset entry.

3.3 DATA RECORDING ENVIRONMENT

The process of recording the dataset was conducted in a study room located on the top floor of a seven-story residential tower. We initiated the recording of a batch of breathing sounds using a Python script to control the microphone, ensuring minimal disturbance to the source and recording equipment.

Once a set of breathing sounds was recorded, we halted the script, adjusted the source-to-device distance, and commenced recording the next category of breathing sounds.

The components of our data acquisition setup included:

(a) **Recording**: Each audio sample was 20 seconds long, saved in the WAV file format, with a python.int16 data type and 16-bit depth.

(b) **Device**: A Dell Inspiron 15 3000 laptop was used for recording, set to a sampling rate of 48,000 Hz and a chunk size of 1,024.

(c) **Setup**: The laptop was prepared to run a Python script capable of capturing acoustic signals through its built-in microphone. It was placed on a stool about 1.5 feet tall, centrally within the room.

(d) **Environment**: The recording was carried out in a study room within a residential building, with a nearly consistent ambient noise level.

(e) **Recording Plan**: Plan for collecting 1,700 samples of 20-second acoustic signals was made. The distance between the respiratory source and the recording device to create a dataset that is both varied and thorough, as detailed in Fig 3.1.

This diligent approach allowed for the methodical collection of data, ensuring a rich and varied dataset for analysis.

CHAPTER 4

PROPOSED METHODOLOGY

In this chapter, we outline our methodology for detecting human respiration sound patterns within inaudible ambient sounds using our custom dataset, **BrthSoundSet**, described in previous chapter.

4.1 PRELIMINARIES

4.1.1 Spectrogram

It is a colored representation of the spectrum of frequencies in a acoustic signal as they evolve over time. It is a two-dimensional plot, where one axis represents time, the other frequency, and the color of each point relates to the amplitude of a particular frequency at a particular time. This multi-dimensional format is quite suitable for the analysis of the acoustic properties, including speech, music, and ambient noises.

4.1.2 Bit Depth

The bit depth equals the number of bits used to represent the amplitude of an acoustic signal. In acoustic signal processing tt is a prime factor in determining the dynamic range and resolution of the audio. Greater dynamic range requires higher bit depth, meaning a larger range of softest to loudest sounds can be captured without distortion from noise.

4.1.3 Deep Learning

Deep learning is a sub-branch of machine learning that uses artificial neural networks, to learn the hidden patterns of data and predict some relevant output. Artificial neural networks are stacked layers of nodes that process and transform data, similar to how organic brain works. Deep learning uses numerous layers of nodes, to learn high-level features and patterns from the input data. Deep learning algorithm has been greatly effective in overcoming long-standing challenges in image recognition, natural language processing and speech recognition.

i) **CNN.** CNN is a type of deep learning algorithm that has its root in electronics stream. Convolution of an input signal with a filter is a traditional and basic step in signal processing. In CNN its adoption for image recognition and processing tasks has displayed unparallel result. It is made up of multiple convolutional layers, pooling layers, and fully connected layers, this trio acting as one set. Convolutional layers smear filters to the input data to extract features, pooling layers reduce the dimensionality of the input, and fully connected layers perform classification or regression tasks.

ii) **BiLSTM.** Sequential data or time-series data are effectively handled by LSTM. A BiLSTM[12], or Bidirectional Long Short-Term Memory, is a type of recurrent neural network (RNN) that can process sequential data in both directions. A BiLSTM consists of two LSTM layers. It captures both past and future context from the data, this technique has proven to be quite effective in enhancing the performance on tasks such as natural language processing, speech recognition, and audio processing.

iii) **Transformer.** A Transformer[13] is a deep learning model that leverages the self-attention mechanism by analyzing the input data by weighting each component individually. It is also used primarily in sequential data like, natural language processing and computer vision with audio processing. This model has an encoder and a decoder as its constituents. The encoder takes the input sequence and maps it into a higher-dimensional space. The decoder takes the encoded intermediate data and generates the output sequence. An essential feature of the Transformer is the attention[14] mechanism, which allows the model to focus on different parts of the input sequences. The attention mechanism computes the relevance of each input element to each output element, and assigns a score to each pair. The higher the score, the more attention the model pays to that pair. The attention mechanism can be applied in different ways, such as self-attention, where the model attends to its own input or output, or cross-attention, where the model attends to both the input and the output.

4.2 DATA PRE-PROCESSING

Initially, audio samples in WAV format are converted into high-resolution spectrogram images of 800x1200 pixels, as shown in fig 6. These spectrograms function as heatmaps with the following components:

i) **X-axis**: Time

ii) **Y-axis**: Frequency

iii) **Pixel Color**: Amplitude, indicating the strength of a frequency at a specific time. After creation, the spectrogram images are carefully cropped to remove any irrelevant portions that do not contribute to the analysis. We have compiled labels in '**labels.csv**' and '**multi_class_labels.csv**' files; the former categorizes recordings by human presence, and the latter includes a column for the distance between the sound source and the recording device.

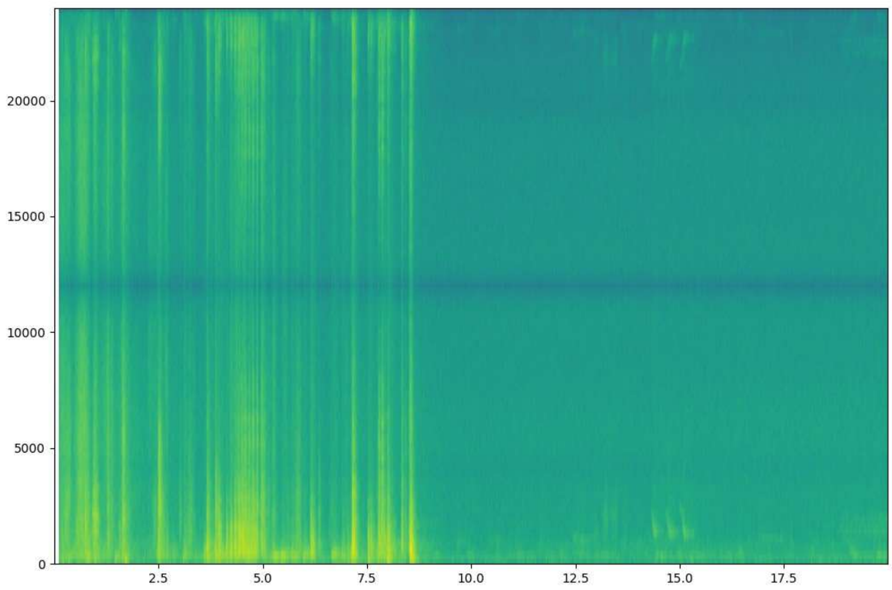

Figure 4.1: An Exemplar of Spectrogram Image

4.2.1 Limitations.

Due to the unique properties of spectrogram images, we avoided standard image augmentation techniques to maintain the integrity of the original orientation and proportions. Therefore, modifications such as cropping, scaling, or rotating were not used. Our final dataset comprises 1700 spectrogram images.

4.2.2 Creation of Labels.

The **BrthSoundSet** Dataset is composed of 1700 audio files in 'wav' format, which have been transformed into an equal number of 'png' spectrogram files. The filenames encode the presence or absence of a human subject and their distance from the recording source, following the conventions outlined in figure 5. Two 'csv' files were generated to serve as labels:

(a) **labels.csv**: This file contains 1000 entries and is utilized for the Binary Classification Task. It includes images from the classes 'No Person', '030', '060', '100', '200', and '300'. In this classification scheme, a '0' denotes the 'No Person' class, indicating the absence of a human, while a '1' signifies the presence of a human.

(b) **multi_class_labels.csv**: This file encompasses all images, each labelled according to its respective category for multi-class classification purposes.

This structured approach to labelling facilitates the accurate categorization of data for different machine learning tasks.

4.3 PROPOSED ARCHITECTURE

Our research utilizes the ResNet18 model, which has been pre-trained on the expansive ImageNet dataset. Pre-training of ResNet18 with million images of 1000 classes, bolster the model with robust feature extraction capabilities. We considered ResNet18 model, which is a convolutional neural network originally designed for optics related tasks as we aim to convert acoustic signals to spectrogram. So, in this scenario, ResNet18 has been repurposed for acoustic classification by coverting audio signals as multi-dimensional visual data in the form of spectrograms. The entire dataset of acoustic signals was first converted into spectrograms. These image spectrograms are then split into three sets: training, validation, and testing, maintaining an 8:1:1 ratio.

4.3.1 Binary Classification

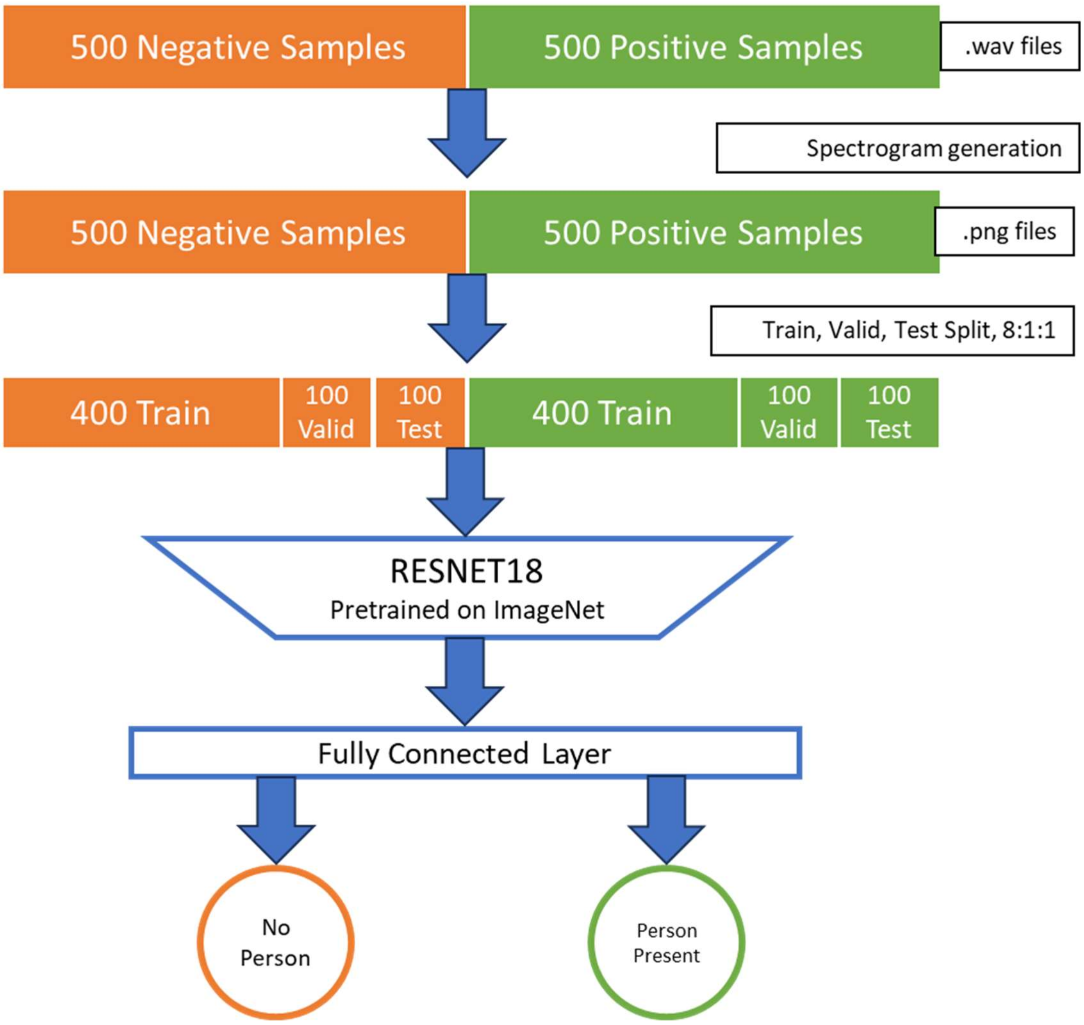

Figure 4.2: Flowchart for Binary Classification Task - Detection of Breathing Sound

We have supplanted this ResNet18 model with fully connected layer to meet our specific training requirements. This layer is crucial as it sums up all the features learned by the model from the spectrograms to give a final prediction. Its output has been fed into sigmoid function that terminates the output into two distinct class labels, facilitating binary classification. This classification process is illustrated in Fig 4.2. This binary classification allows the model to determine whether a human is present in the acoustic recording based on the spectrogram.

4.3.2 Multi-Class Classification

Furthermore, we have modified the model to predict the distance between the sound source and the recording device by classifying the audio signals into 13 distinct classes. For this purpose, we reconfigured the last layer to perform multi-class classification, as demonstrated in Fig 4.3. The fully connected layer serves as the decision-making component of the network, where the combination of extracted features leads to the final classification. The model outputs a classification into one of the 13 predefined classes based on the patterns recognized in the spectrogram images.

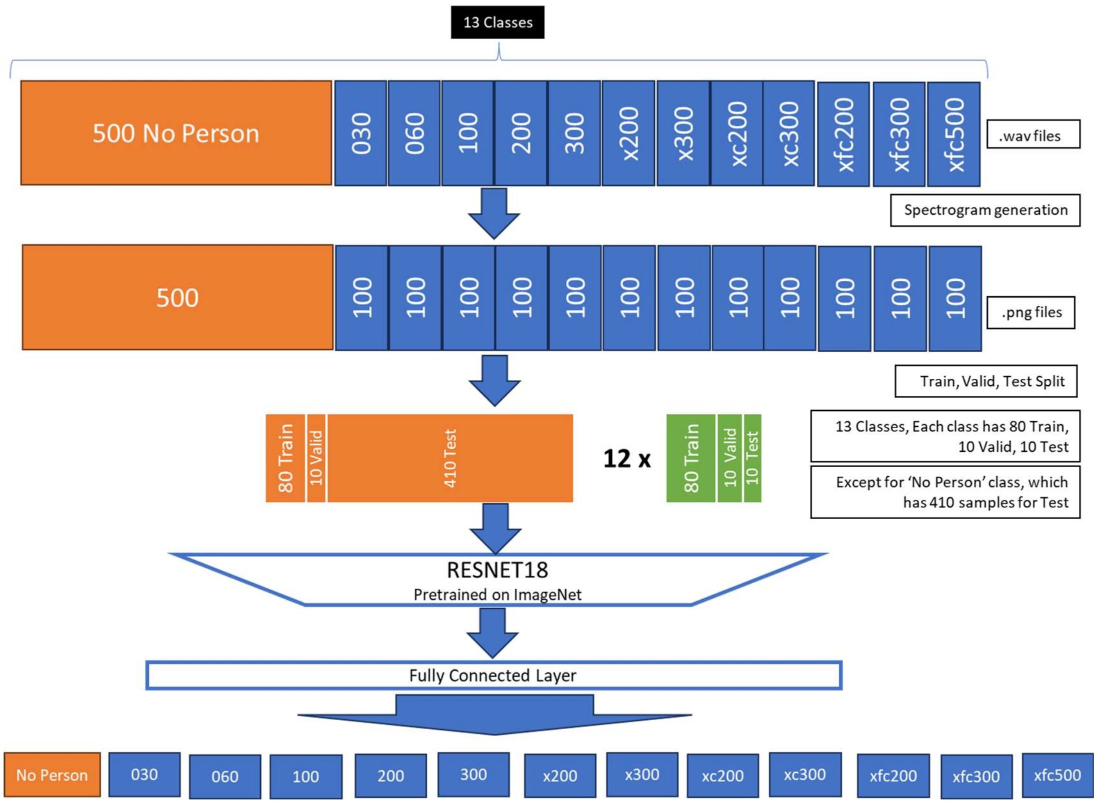

Figure 4.3: Flowchart for Multi-Class Classification - Estimating Distance

The ResNet18 architecture is notable for its skip connections[35], which support identity mapping and merge their outputs with those of subsequent layers, thereby enhancing the overall learning process.

For our study, input images were resized to 512 x 748 pixels prior to processing via the ResNet18 model. The dataset, consisting of 1700 specimens, was apportioned into training, validation, and testing subsets in an 8:1:1 ratio. Training proceeded with a batch size of 10. The use of spectrograms bridges the gap between audio and visual data, allowing the RESNET18 model to apply its image classification strengths to the domain of audio analysis.

4.3 HYPERPARAMETERS

In the machine learning component of our study, we configured a set of hyperparameters to optimize the training process. In the training process we used a batch of 10 samples per epoch. We chose Stochastic Gradient Descent (SGD) as optimiser, which incorporated a momentum of 0.9 and a weight decay factor of 0.0005 to prevent overfitting. The learning rate was set at 0.0002 to ensure gradual and steady model updates. To adjust the learning rate during training, we used a StepLR[36] scheduler, which modifies the learning rate after a certain number of epochs. Specifically, the learning rate was scheduled to decay by a factor of 0.9 every 4 epochs.

4.4 ARCHITECTURE OF PROPOSED MODEL

The proposed architecture of this enhanced ResNet18 based model is as shown below in Fig 4.4.

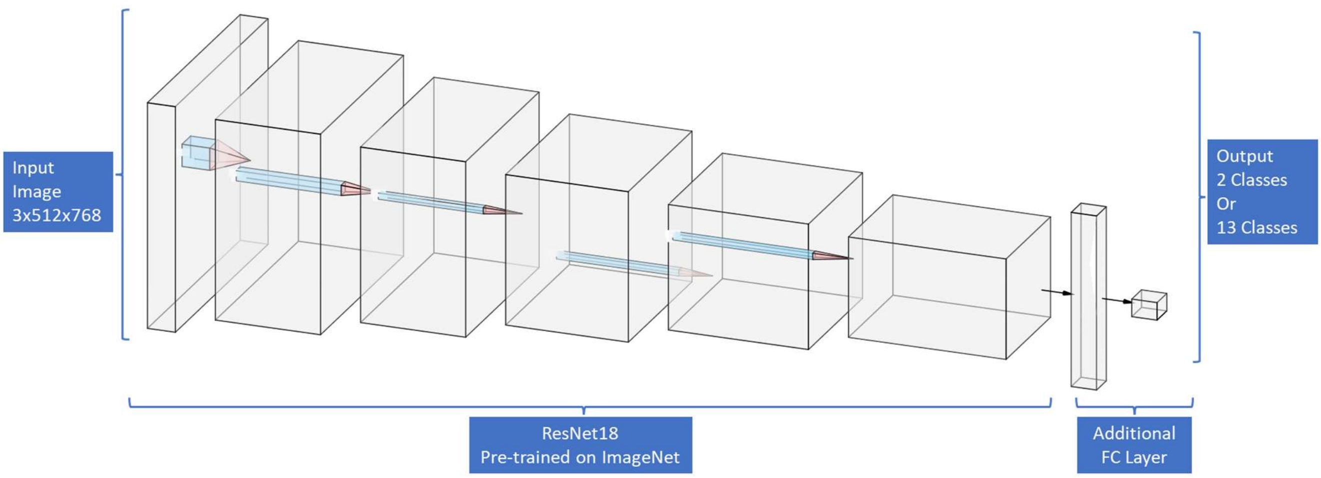

Figure 4.4: Architecture of Proposed Model

4.5 ADVANTAGE OF PROPOSED SYSTEM

The use of spectrograms bridges the gap between acoustic and visual data. This creative approach to audio signal classification by leveraging the visual pattern recognition capabilities of a pre-trained deep learning model is novelty in this work. Allowing the ResNet18 model to apply its image classification strengths to the domain of audio analysis has displayed high accuracy. The advantages of this approach for acoustic signal classification, are multifaceted:

(a) **Leveraging Pretraining**: By utilizing ResNet18, which is pretrained on ImageNet, allows the model to swiftly recognize patterns in spectrograms, which are essentially visual representations of audio signals.

(b) **Systematic Data Processing Pipeline**: The ingenious approach to data preparation, including the conversion of '.wav' files to spectrograms and then to '.png' files, followed by dataset splitting, ensures a systematic pipeline that facilitates the model's training and evaluation.

(c) **Transfer Learning**: The use of transfer learning not only fastens the training process but also enhances the model's performance, especially when the amount of available training data is limited.

(d) **Robustness to Environment**: By including samples recorded under various conditions (Line of Sight and Non-Line of Sight), the model is likely to be more robust to environmental vagaries, which is crucial for real-world applications.

CHAPTER 5

EXPERIMENTAL SETUP & RESULT ANALYSIS

5.1 EXPERIMENTAL SETUP

In this sub-section, the experimental setup for the model is diligently detailed, focusing on the software and hardware components essential for its operation.

5.1.1 Software Requirements

i) **Platform**: The model operates within the Google Colab environment, a cloud-based platform that provides a seamless interface for executing Python code.

ii) **APIs and Drivers**: Essential APIs and drivers are required to ensure compatibility and functionality within the Google Colab ecosystem.

iii) **Software**: Google Colab is the cloud-based software platform used due to its accessibility and integration with various Google services.

iv) **Language**: Python 3.9 programming language was used for its widespread use and support within the Artificial Intelligence domain.

5.1.2 Hardware Requirements

i) **Processing Power**: We have used the Google Colab's Tensor Processing Units (TPU) for training model. It has offered high-speed computation capabilities essential for processing large datasets and complex algorithms.

ii) **Memory**: Memory allocation is managed on demand by Google Colab, allowing for flexible scaling based on the model's requirements.

iii) **Secondary Storage**: Any Secondary storage is suitable that provides adequate data transfer speed for image transfers.

5.1.3 Libraries/Packages

i) **Pytorch**: An open-source machine learning library favoured for its ease of use and efficiency in creating and training neural networks.

ii) **PyAudio**: A set of Python bindings for PortAudio, a cross-platform audio I/O library, enabling the model to record and play audio.

iii) **Numpy**: We have used Numpy mathematical package for scientific computing in Python. It provides support for large, multi-dimensional arrays and matrices.

iv) **Pandas**: For data structures and data analysis tools we have used Pandas python packages, which is, suitable for manipulating numerical tables and time series.

v) **Matplotlib**: For visualising our result and spectrogram we has used Matplotlib library for Python.

Above mentioned comprehensive setup of hardware and software has supported model training and testing, so that it is well-equipped to handle the computational demands of detecting human breathing patterns through acoustic signals. The combination of Google Colab's cloud-based resources with the Python and its associated libraries provided a robust framework for conducting this innovative research.

5.2 DATASET DESCRIPTION

The '**BrthSoundSet'** dataset is a comprehensive collection of audio samples designed for the purpose of training deep learning models to recognize and classify breathing sounds.

(a) **Total Samples**: The dataset consists of 1700 audio files in '.wav' format, which have been converted into spectrogram images as '.png' files.

(b)**Encoding**: The presence or absence of a human subject, as well as their distance from the recording source, is encoded in the filenames according to the nomenclature outlined in Fig 3.1

5.3 PERFORMANCE EVALUATION MATRIX

The evaluation of our model performance encompasses various metrics such as precision, recall, accuracy and F1 score. These metrics are defined as follows:

5.3.1 Confusion Matrix. The Confusion Matrix, as delineated in Fig 5.1 and Fig 5.2, serves as a tabular juxtaposition of Actual versus Predicted classifications. The Y-axis enumerates the true labels of the data samples, with '0' denoting the absence and '1' the presence of a person within the room. Conversely, the X-axis is annotated with the labels predicted by the model. For multi-class confusion matrix, correlation between label and class can be read out from Table 5.1.

Table 5.1 – Labels vs Class used in Multi-Class Classification Model

Label	Class	Label	Class
0	No Person	**7**	x300 cm
1	100 cm	**8**	xc200 cm
2	200 cm	**9**	xc300 cm
3	030 cm	**10**	xfc200 cm
4	300 cm	**11**	xfc300 cm
5	060 cm	**12**	xfc500 cm
6	x200 cm		

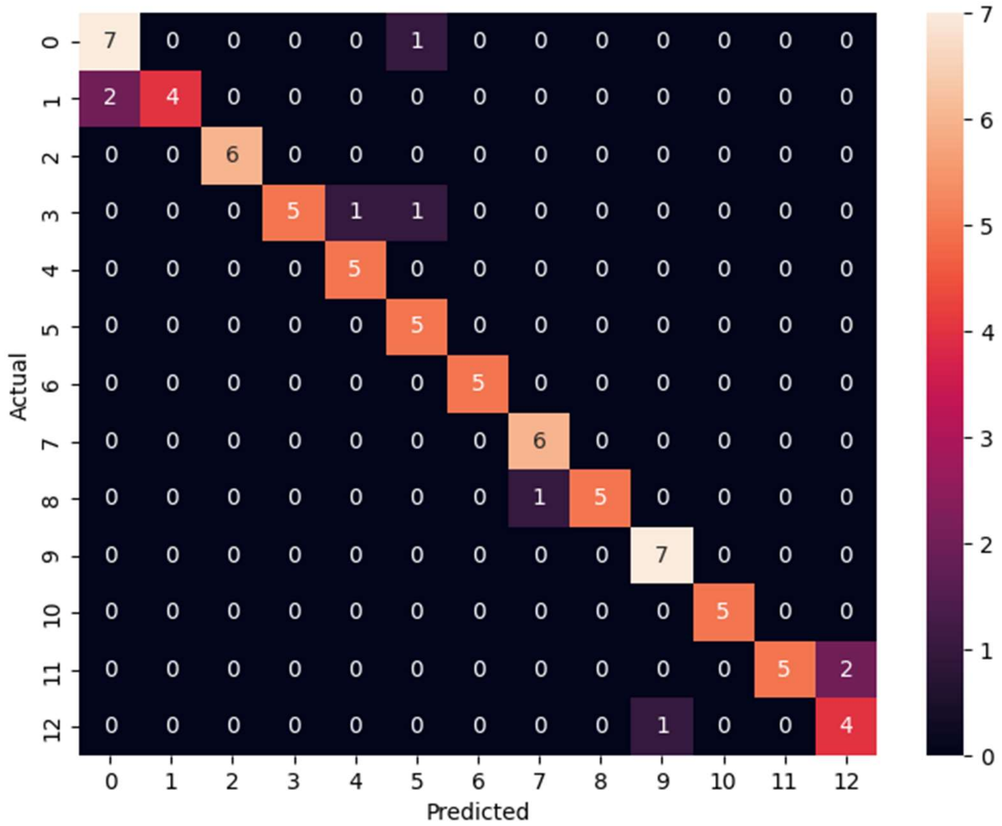

Figure 5.15: Confusion Matrix for Multi-Class Classification

Actual/ Predicted	No Person ‘0’	Person Present ‘1’
No Person ‘0’	43	7
Person Present ‘1’	0	50

Figure 5.2: Confusion Matrix for Binary Classification Task

5.3.2 Precision: This metric quantifies the proportion of instances where the presence of a person is correctly identified by the model. It is computed as

$$\text{Precision} = TP/TP+FP \qquad (2)$$

where (TP) represents true positives and (FP) false positives.

5.3.3 Recall/Sensitivity: This measure reflects the ratio of correctly identified presence samples to the total number of actual present. It is calculated as

$$\text{Recall} = TP/TP+FN \tag{3}$$

with (FN) denoting false negatives.

5.3.4 Specificity: This parameter represents the fraction of correctly predicted absence samples relative to the aggregate of absence samples. It is defined as

$$\text{Specificity}= TN/TN+FP \tag{4}$$

5.3.5 Accuracy: This statistic conveys the proportion of accurate predictions made by the model out of all predictions. It is expressed as

$$\text{Accuracy}=(TP+TN)/\text{Total} \tag{5}$$

5.3.6 F1 Score: The F1 score is the harmonic mean of precision and recall, providing a balance between the two metrics. It is calculated as:

$$\text{F1 Score} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \tag{6}$$

5.4 RESULT ANALYSIS

The model exhibited commendable efficacy, as evidenced by a 93% and 87% accuracy rate in binary classification (detection of human breath) and multi-class classification (predicting distance between source and recording device) as shown in Table 5.2. The challenge of discerning human respiration within a span of up to 500 cm was adeptly transmuted into a multi class classification task, which the model proficiently navigated.

Table 5.2. Results for Binary and Multi-Class Classification

	Binary Classification	Multi-Class Classification
F1 Score	0.93	0.884
Sensitivity	0.93	0.885
Precision	0.939	0.905
Accuracy	0.93	0.885

5.5 DISCUSSION & FINDINGS

A noteworthy observation is the model's swift attainment of over 80% accuracy in the preliminary epochs as shown in Fig 5.3 for breathing detection task, with this metric escalating to near-perfection within 8-9 epochs, indicative of the model's adeptness at internalizing pivotal patterns within the spectrogram images.

5.5.1 Breathing Detection Task

The model demonstrated rapid proficiency, achieving over 80% accuracy in the initial epochs, as depicted in Fig 5.3. This swift rise to near-perfect accuracy within just 8-9 epochs underscores the model's capability to discern critical patterns in the spectrogram images. The impressive F1 score and accuracy are attributable to the fact that, while the natural frequency of human respiration falls below 1 Hz, the spectrograms encapsulate the entire frequency range. As a result, the model's analytical focus is concentrated on a discrete, localized region situated in the lower-left quadrant of each spectrogram, leading to precise breathing detection.

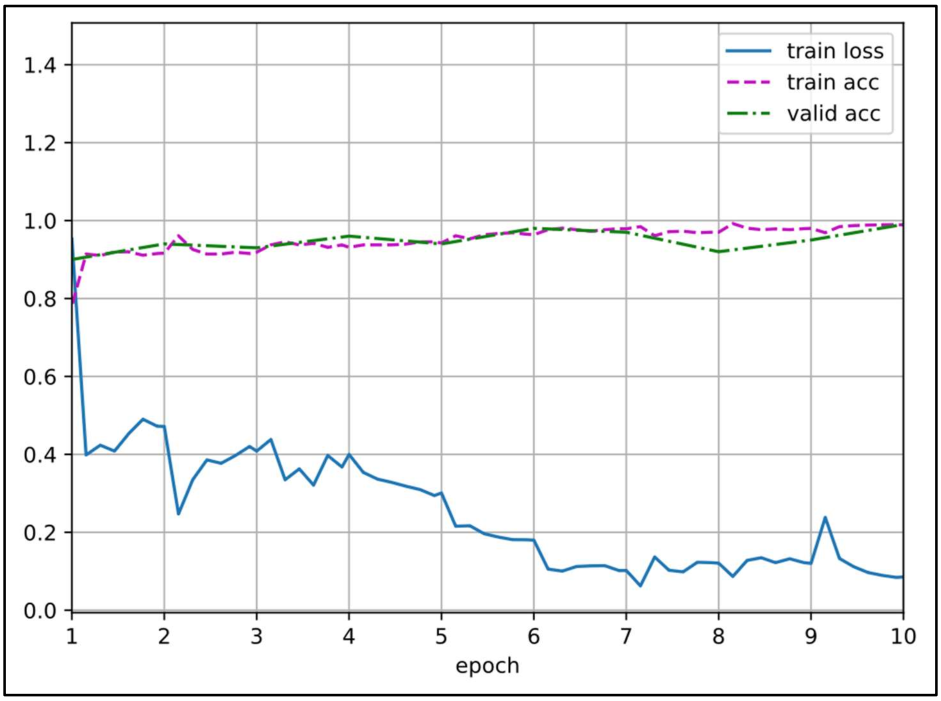

Figure 5.36: Breathing Detection Task – Loss & Accuracy

5.5.2 Estimation of Distance (Multi-class Classification Task)

During the model's training as shown in Fig 5.4, two principal trends were observed:

(a) **Gradual Improvement in Training Accuracy**: The model's training accuracy exhibited a steady increase with each epoch. This gradual enhancement signifies the model's ongoing learning from the training data, fine-tuning its internal parameters to more accurately reflect the data's inherent patterns.

(b) **Consistent Reduction in Training Loss**: Alongside the accuracy gains, there was a consistent reduction in training loss across epochs. This decrease in loss, a quantification of the model's prediction error, indicates an approach toward the actual target values. The initial sharp decline, followed by a levelling off, suggests that the model rapidly assimilates substantial information at the outset. However, the incremental learning from each subsequent epoch appears to taper off.

These observations suggest that the model is effectively learning and enhancing its performance over time. Nevertheless, the levelling off of both accuracy and loss points to a potential saturation in learning from the provided dataset. Beyond a certain juncture, additional training epochs do not seem to yield further improvements in performance.

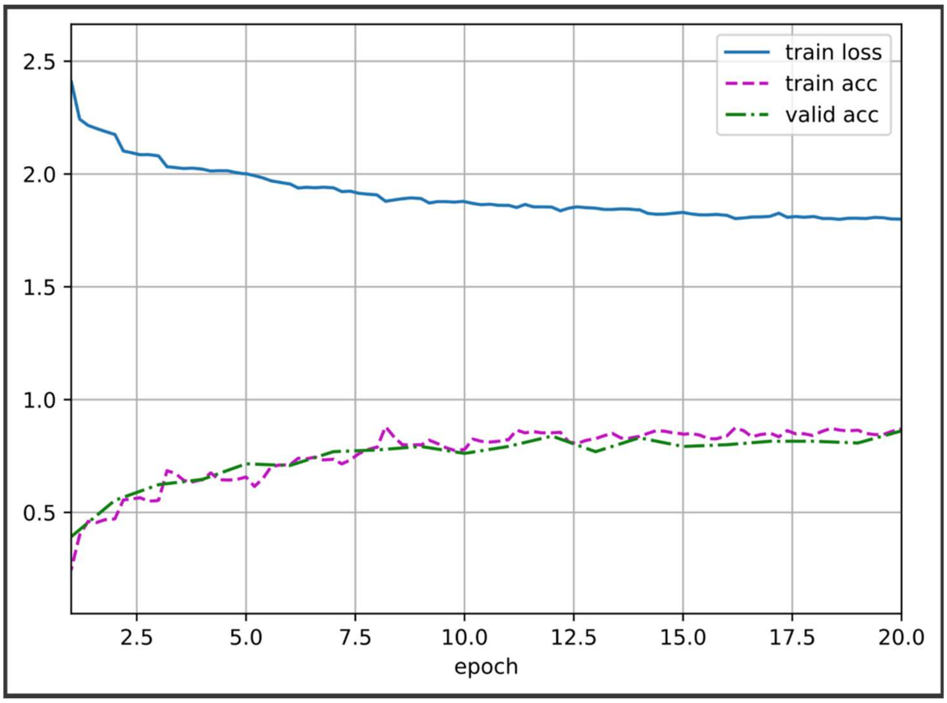

Figure 5.4: Distance Estimation Task – Loss & Accuracy

CHAPTER 6

CONCLUSION, FUTURE WORK & SOCIAL IMPACT

6.1 CONCLUSION

The conclusion of this thesis underscores the significant strides made in the realm of concealed human detection; a domain fraught with challenges for security personnel. Traditional methods such as through-the-wall radar and WiFi antenna sensing, while effective in NLOS conditions, are hindered by their cumbersome nature and dependency on existing infrastructure. In contrast, the sound-based detection system proposed in this paper is passive, unobtrusive, and capable of operating effectively in NLOS scenarios.

This research marks the first endeavour to design a deep learning model tailored specifically for human breathing detection using sound. A systematic review of a broad range of techniques, was undertaken, which guided this novel approach and precipitated in to successful identification of human presence. The novelty of this task lies in its unique approach, as no precedent exists for such an application.

The success of this innovative approach has been buttressed through the use of deep learning i.e, pre-trained ResNet18 architecture, which has collectively enabled the model to overcome the constraints of distance and noise. The accuracy achieved in detecting human breathing patterns from up to five meters away is a confirmation to the model's efficacy.

The creation and utilization of the '**BrthSoundSet'** dataset have been quint-essential in development of the model. This highlights the importance of specialized data for such niche problems. The findings of this study provide a potential solution for enhanced human detection in conditions where visibility is severely compromised, by providing valuable insights to for disaster response and security monitoring.

6.2 LIMITATIONS

The current model's performance and robustness is limited by the dataset and the architecture of the model. The extensive diversity of acoustic environments found in nature, has not been capture in the current dataset, which is crucial for training a model to operate under various real-world conditions. While the current dataset was created a pilot, it is sufficient to test the initial hypothesis, however it does not encompass the full spectrum of acoustic variations that the model may encounter. Few derived limitations are as follows:

(a) **Environmental Noise**: The model's performance could be affected by unseen ambient noise, which has not been adequately captured in the training dataset.

(b) **Microphone Sensitivity**: The microphones used for recording can significantly affect the quality of acoustic signals and, consequently, the model's performance.

(c) **Dynamic Environments**: The dynamic environments where the acoustic properties changes over time, such as in moving vehicles or crowded areas will require a dataset that captures all such variations.

(d) **Dependence on Visual Representation**: Spectrogram images generated from acoustic signals are the bottleneck in model's efficiency. Any random inaccuracies in the visual representation of audio signals could potentially lead to misclassification.

(e) **Real-Time Processing**: For real-time acoustic signal processing, this model is unsuitable as it requires significant time for spectrogram generation and model inference.

Any future work shall aim to address these limitations by expanding the dataset, improving noise reduction techniques, utilising microphone arrays, and refining the deep learning model to better generalize across different scenarios. Integrated testing and validation in diverse settings will be essential to evolve the model into a robust tool for human detection using sound analysis.

6.3 SOCIAL & INDUSTRIAL APPLICATION

The social and industrial applications of this thesis work on human detection through sound analysis are multifarious and have the potential to impact various sectors significantly.

6.3.1 Social Applications:

(a) **Disaster Relief**: Following natural disasters like earthquakes or building collapses, this approach can be used to detect survivors buried beneath debris, allowing prompt rescue efforts and perhaps saving lives.

(b) **Healthcare Management**: The non-invasive nature of sound-based detection will be beneficial for remote health monitoring, particularly for those with respiratory illnesses. This will enhance the quality of healthcare that is provided.

(c) **Enhanced Home Security**: Without the need for cameras, homeowners can use this to detect intrusions or the presence of people in restricted areas, protecting their privacy.

(d) **Wildlife Research and Conservation**: By using this technology, wildlife researchers may monitor the existence of animals in their natural habitats without upsetting them, which will help with their rehabilitation and conservation efforts.

6.3.2 Industrial Applications:

(a) **Remote Surveillance**: This model can be used for surveillance in locations with high levels of security, such as airports. Detecting human presence in restricted region without the requirement for visual confirmation.

(b) **Intelligent Buildings**: By incorporating such a model into building systems, security and safety procedures can be improved. This is because human presence can be detected in the event of an emergency, such as a fire, in low-visibility situations.

(c) **Counterterrorism Operations**: This technology can be used by the military for intelligence gathering and reconnaissance, identifying the presence of enemies in veiled areas and in close quarters combat situations.

(d) **Search and Rescue**: In dangerous areas, AI-based robots equipped with this detecting technology may locate people by listening to their breathing noises.

The capacity of the deep learning model to differentiate breathing sounds from background noise creates new opportunities for non-intrusive detection and monitoring in both industrial and social environments. The thesis study provides a novel technique to human identification through sound analysis, laying the foundation for these applications. The use of these models will produce workable solutions as they are improved and strengthened.

6.4 FUTURE WORK

The future work stemming from this thesis and its finding will aim at enhancing the capabilities and applications of acoustic-based detection systems. The following objectives outline the forecasted research areas:

(a) **Dataset Expansion**: It’s important to create and annotate an extensive and varied dataset of acoustic signals, capturing the myriad acoustic environments encountered in urban and rural settings, both indoors and outdoors, and ranging from vibrant to serene scenarios. It will make future model more robust.

(b) **Architectural Evaluation**: A comparative analyses of various deep learning architectures, including RNNs, and Attention mechanisms and subsequently assessing their efficacy in the context of acoustic-based human detection.

(c) **Factor Analysis**: Its essential to investigate the effect of diverse factors such as ambient noise levels, proximity of multiple subjects, and physical obstructions, wind flow, on the precision and dependability of the sound-based detection system.

(d) **Device Development**: An essential step would be to design and create a portable, user-friendly device capable of real-time implementation of the sound-based detection system, contributing actionable insights to security forces.

This pioneering work not only contributes to the body of knowledge on human detection technologies but also sets the stage for advanced studies into the integration of deep learning with ambient sound analysis. The anticipated adoption of this technology in various sectors signifies a major leap forward in safety protocols and the improvement of critical outcomes during urgent situations.

REFERENCES

[1] H. Ferreira-Cardoso *et al.*, "Comparison of lung auscultation between smartphone and digital stethoscope in patients with asthma: a feasibility study," in *01.04 - M-Health/e-health*, European Respiratory Society, Sep. 2022, p. 1859. doi: 10.1183/13993003.congress-2022.1859.

[2] A. S. Bugaev, I. A. Vasil'ev, S. I. Ivashov, and V. V. Chapurskii, "Radar methods of detection of human breathing and heartbeat," *Journal of Communications Technology and Electronics*, vol. 51, no. 10, pp. 1154–1168, Oct. 2006, doi: 10.1134/S1064226906100056.

[3] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/NATURE14539.

[4] M. N. Markandeya and U. R. Abeyratne, "Smart Phone Based Snoring Sound Analysis to Identify Upper Airway Obstructions," *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany*, pp. 4233–4236, 2019.

[5] J. Eargle, "Chapter 2 - Basic Sound Transmission and Operational Forces on Microphones," in *The Microphone Book (Second Edition)*, Second Edition., J. Eargle, Ed., Oxford: Focal Press, 2004, pp. 7–21. doi: https://doi.org/10.1016/B978-0-240-51961-6.50006-7.

[6] B. Islam *et al.*, "BreathTrack," *Proc ACM Interact Mob Wearable Ubiquitous Technol*, vol. 5, no. 3, pp. 1–22, Sep. 2021, doi: 10.1145/3478123.

[7] N. Sharma *et al.*, "Coswara -- A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis," May 2020, doi: 10.21437/Interspeech.2020-2768.

[8] J. Eargle, "Chapter 1 - A Short History of the Microphone," in *The Microphone Book (Second Edition)*, Second Edition., J. Eargle, Ed., Oxford: Focal Press, 2004, pp. 1–6. doi: https://doi.org/10.1016/B978-0-240-51961-6.50005-5.

[9] J. Eargle, "Chapter 4 - The Pressure Gradient Microphone," in *The Microphone Book (Second Edition)*, Second Edition., J. Eargle, Ed., Oxford: Focal Press, 2004, pp. 50–65. doi: https://doi.org/10.1016/B978-0-240-51961-6.50008-0.

[10] E. Kurz, F. Pfahler, and M. Frank, "Comparison of first-order Ambisonics microphone arrays," in *3rd International Conference on Spatial Audio, ICSA*, 2015.

[11] E. O. Brigham and R. E. Morrow, "The fast Fourier transform," *IEEE Spectr*, vol. 4, no. 12, pp. 63–70, Dec. 1967, doi: 10.1109/MSPEC.1967.5217220.

[12] Y. Zhao, R. Yang, G. Chevalier, and M. Gong, "Deep Residual Bidir-LSTM for Human Activity Recognition Using Wearable Sensors," Aug. 2017.

[13] T. Lin, Y. Wang, X. Liu, and X. Qiu, “A Survey of Transformers,” Jun. 2021, [Online]. Available: http://arxiv.org/abs/2106.04554

[14] A. Vaswani *et al.*, “Attention Is All You Need,” Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.03762

[15] S. Costanzo, “Software-Defined Doppler Radar Sensor for Human Breathing Detection,” *Sensors*, vol. 19, no. 14, p. 3085, Jul. 2019, doi: 10.3390/s19143085.

[16] A. Dell’aversano, A. Natale, A. Buonanno, and R. Solimene, “Through the Wall Breathing Detection by Means of a Doppler Radar and MUSIC Algorithm,” vol. 1, no. 3, 2017, doi: 10.1109/LSENS.2017.2704902.

[17] D. Exploring Radar *et al.*, “Citation: Exploring Radar Micro-Doppler Signatures for Recognition of Drone Types,” 2023, doi: 10.3390/drones7040280.

[18] E. Cardillo and A. Caddemi, “Radar Range-Breathing Separation for the Automatic Detection of Humans in Cluttered Environments,” *IEEE Sens J*, vol. 21, no. 13, 2021, doi: 10.1109/JSEN.2020.3024961.

[19] G. Li, Y. Ge, Y. Wang, Q. Chen, and G. Wang, “Detection of Human Breathing in Non-Line-of-Sight Region by Using mmWave FMCW Radar,” *IEEE Trans Instrum Meas*, vol. 71, pp. 1–11, 2022, doi: 10.1109/TIM.2022.3208266.

[20] M. Husaini *et al.*, “Non-Contact Breathing Monitoring Using Sleep Breathing Detection Algorithm (SBDA) Based on UWB Radar Sensors,” 2022, doi: 10.3390/s22145249.

[21] C. Chen *et al.*, “TR-BREATH: Time-Reversal Breathing Rate Estimation and Detection,” *IEEE Trans Biomed Eng*, vol. 65, no. 3, 2018, doi: 10.1109/TBME.2017.2699422.

[22] L. Guan *et al.*, “Multi-Person Breathing Detection With Switching Antenna Array Based on WiFi Signal,” *IEEE J Transl Eng Health Med*, vol. 11, pp. 23–31, 2023, doi: 10.1109/JTEHM.2022.3218638.

[23] S. Bae *et al.*, “Prospective validation of smartphone-based heart rate and respiratory rate measurement algorithms,” *Communications Medicine*, vol. 2, no. 1, p. 40, Apr. 2022, doi: 10.1038/s43856-022-00102-x.

[24] D. Tran-Anh, N. H. Vu, K. Nguyen-Trong, and C. Pham, “Multi-task learning neural networks for breath sound detection and classification in pervasive healthcare,” *Pervasive Mob Comput*, vol. 86, p. 101685, Oct. 2022, doi: 10.1016/j.pmcj.2022.101685.

[25] J. Hong *et al.*, “End-to-End Sleep Staging Using Nocturnal Sounds from Microphone Chips for Mobile Devices,” *Nat Sci Sleep*, vol. Volume 14, pp. 1187–1201, Jun. 2022, doi: 10.2147/NSS.S361270.

[26] Y. Nam, B. A. Reyes, and K. H. Chon, "Estimation of Respiratory Rates Using the Built-in Microphone of a Smartphone or Headset," *IEEE J Biomed Health Inform*, vol. 20, no. 6, p. 1493, 2016, doi: 10.1109/JBHI.2015.2480838.

[27] L. Wang and Z. Jiang, "Tidal Volume Level Estimation Using Respiratory Sounds," 2023, doi: 10.1155/2023/4994668.

[28] A. Semmad and M. Bahoura, "Scalable serial hardware architecture of multilayer perceptron neural network for automatic wheezing detection," *Microprocess Microsyst*, vol. 99, Jun. 2023, doi: 10.1016/j.micpro.2023.104844.

[29] T. Rahman *et al.*, "QUCoughScope: An Intelligent Application to Detect COVID-19 Patients Using Cough and Breath Sounds," *Diagnostics*, vol. 12, no. 4, Apr. 2022, doi: 10.3390/diagnostics12040920.

[30] Daniel, T. W. Cenggoro, and B. Pardamean, "A systematic literature review of machine learning application in COVID-19 medical image classification," *Procedia Comput Sci*, vol. 216, pp. 749–756, Jan. 2023, doi: 10.1016/J.PROCS.2022.12.192.

[31] F.-S. Hsu *et al.*, "Benchmarking of eight recurrent neural network variants for breath phase and adventitious sound detection on a self-developed open-access lung sound database—HF_Lung_V1," *PLoS One*, vol. 16, no. 7, p. e0254134, Jul. 2021, doi: 10.1371/journal.pone.0254134.

[32] P. Corbishley and E. Rodríguez-Villegas, "Breathing Detection: Towards a Miniaturized, Wearable, Battery-Operated Monitoring System," *IEEE Trans Biomed Eng*, vol. 55, no. 1, 2008, doi: 10.1109/TBME.2007.910679.

[33] P. Y. Chan, N. P. Ryan, D. Chen, J. McNeil, and I. Hopper, "Novel wearable and contactless heart rate, respiratory rate, and oxygen saturation monitoring devices: a systematic review and meta-analysis," *Anaesthesia*, vol. 77, no. 11, pp. 1268–1280, Nov. 2022, doi: 10.1111/anae.15834.

[34] R. K. Singh and R. Katarya, "Recent Trends in Human Breathing Detection Using Radar, WiFi and Acoustics," in *2023 6th International Conference on Recent Trends in Advance Computing (ICRTAC)*, IEEE, Dec. 2023, pp. 530–536. doi: 10.1109/ICRTAC59277.2023.10480776.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015.

[36] I. Loshchilov and F. Hutter, "SGDR: Stochastic Gradient Descent with Warm Restarts," Aug. 2016.

[37] "Polysomnography(PSG)," https://www.psychdb.com/_media/sleep/1-introduction/sleep-graph.png.

[38] M. Alkhodari and A. H. Khandoker, “Detection of COVID-19 in smartphone-based breathing recordings: A pre-screening deep learning tool,” *PLoS One*, vol. 17, no. 1, p. e0262448, Jan. 2022, doi: 10.1371/journal.pone.0262448.

[39] M. G. Campana, A. Rovati, F. Delmastro, and E. Pagani, “L ^{3} -Net Deep Audio Embeddings to Improve COVID-19 Detection from Smartphone Data,” in *2022 IEEE International Conference on Smart Computing (SMARTCOMP)*, IEEE, Jun. 2022, pp. 100–107. doi: 10.1109/SMARTCOMP55677.2022.00029.

[40] S. Yu, Y. Ding, K. Qian, B. Hu, W. Li, and B. W. Schuller, “A Glance-and-Gaze Network for Respiratory Sound Classification,” in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, May 2022, pp. 9007–9011. doi: 10.1109/ICASSP43922.2022.9746053.

[41] L. Kranthi Kumar and P. J. A Alphonse, “THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS COVID-19: respiratory disease diagnosis with regularized deep convolutional neural network using human respiratory sounds,” *Eur Phys J Spec Top*, vol. 231, pp. 3673–3696, 2022, doi: 10.1140/epjs/s11734-022-00649-9.

[42] G. Furman *et al.*, “Remote Analysis of Respiratory Sounds in Patients With COVID-19: Development of Fast Fourier Transform–Based Computer-Assisted Diagnostic Methods,” *JMIR Form Res*, vol. 6, no. 7, p. e31200, Jul. 2022, doi: 10.2196/31200.

[43] V. S. Fahed, E. P. Doheny, and M. M. Lowery, “Random Forest Classification of Breathing Phases from Audio Signals Recorded using Mobile Devices,” in *INTERSPEECH 2023*, ISCA: ISCA, Aug. 2023, pp. 2738–2742. doi: 10.21437/Interspeech.2023-2089.

[44] E. P. Doheny *et al.*, “Estimation of respiratory rate and exhale duration using audio signals recorded by smartphone microphones,” *Biomed Signal Process Control*, vol. 80, p. 104318, Feb. 2023, doi: 10.1016/j.bspc.2022.104318.

[45] H. H. Tran *et al.*, “Prediction of Sleep Stages Via Deep Learning Using Smartphone Audio Recordings in Home Environments: Model Development and Validation,” *J Med Internet Res*, vol. 25, p. e46216, Jun. 2023, doi: 10.2196/46216.

[46] Interesting Engineering, “The Doppler effect and its application in real life,” https://inteng-storage.s3.amazonaws.com/img/iea/qrwBnNJpw5/sizes/doppler-shift-light_resize_md.jpg.

[47] W. Liu, S. Chang, F. Li, Y. Xu, S. Yan, and Y. Liu, “Contactless Breathing Airflow Detection on Smartphone,” *IEEE Internet Things J*, vol. 10, no. 4, 2023, doi: 10.1007/978-3-030-85928.

[48] R. Phoophuangpairoj, “Recognizing Breathing Sounds Using HMMs and Grammar,” in *2020 - 5th International Conference on Information Technology (InCIT)*, IEEE, Oct. 2020, pp. 144–149. doi: 10.1109/InCIT50588.2020.9310966.

[49] W. Li, B. Tan, and R. J. Piechocki, "Non-contact breathing detection using passive radar," in *2016 IEEE International Conference on Communications (ICC)*, IEEE, May 2016, pp. 1–6. doi: 10.1109/ICC.2016.7511389.

LIST OF PUBLICATIONS/ACCEPTANCES & THEIR PROOFS

1. R. K. Singh and R. Katarya, "**Recent Trends in Human Breathing Detection Using Radar, WiFi and Acoustics**" in *2023 6th International Conference on Recent Trends in Advance Computing (ICRTAC)*, IEEE, Dec. 2023, pp. 530–536. doi: 10.1109/ICRTAC59277.2023.10480776.

ICRTAC-2023

VIT Vellore Institute of Technology (Deemed to be University under section 3 of UGC Act, 1956) CHENNAI

IEEE Madras Section

Certificate of Appreciation

PaperID: 314

This is to certify that Dr./Prof./Mr./Ms.. Ravi Kant Singh of Delhi Technological University,Delhi has authored a paper titled Recent Trends in Human Breathing Detection Using Radar, WiFi and Acoustics in the 6th "International Conference on Recent Trends in Advanced Computing Theme : Cutting Edge Technologies in Computing – SDG (ICRTAC 2023)" organized by the School of Computer Science and Engineering (SCOPE) in Collaboration with VIT Business School from December 14th to 15th, 2023 at Vellore Institute of Technology, Chennai.

Dr. Thomas Abraham J V, Conference Chair

Dr. Vedhapriyavadhana R, Conference Chair

Dr. Ganesan R, Dean-SCOPE

VIT - A place to learn; A chance to grow

2. Paper title: '**Spectro-Acoustic Breath Detection – A ResNet18 Based Approach**' is accepted in '2nd International Conferences on Optimization Techniques in Engineering and Technology (ICOTET 2024)' SCIE in May 2024.

Acceptance : ICOTET 2024 Inbox

ICOTET2024 <icotetdgi2024@gmail.com> 3:06 PM (4 hours ago)
to me, icotet

Greetings from ICOTET 2024!

Dear Author (s)

We are pleased to inform you that **Paper ID 2502** entitled " Spectro-Acoustic Breath Detection: A ResNet18-Based Approach" submitted by you has been accepted by the 2nd International Conference on Optimization Techniques in Engineering and Technology Engineering (ICOTET 2024).

You are advised to register for the conference by 16th of May, 2024! Payment details for registration can be found at the bottom of this email.

You are requested to fill out the following Google form for the registration and payment information etc.:

https://forms.gle/mqyRFhx45hqkJ6cx7

All the registered and presented papers for the 2nd ICOTET 2024 will be published in the AIP Conference Proceedings (Scopus Index) and Springer Nature Conference Proceedings (Scopus Index). Please note that the plagiarism level of the paper should not exceed 15%.

For further details, please visit the official website: https://www.icotet.in/registration

Thanks & Regards

Organizing Committee

ICOTET 2024.

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis: **A Novel Approach for Breathing Sound Detection System using Deep Learning**.

Total Pages **54**

Name of the Scholar: **Ravi Kant Singh, (2K22/AFI/30)**

Supervisor: **Prof Rahul Katarya**

Department: **Department of Computer Science and Engineering**

This is to report that the above thesis was scanned for similarity detection. Process and outcome are given below:

Software used: **Turnitin** Similarity Index: **7%** Total Word Count: **12023**

Date

Candidate's Signature **Signature of Supervisor**

PAPER NAME

Only_Contect MTech Thesis.pdf

WORD COUNT	CHARACTER COUNT
12023 Words	**66874 Characters**
PAGE COUNT	FILE SIZE
54 Pages	**2.1MB**
SUBMISSION DATE	REPORT DATE
May 25, 2024 12:00 PM GMT+5:30	**May 25, 2024 12:01 PM GMT+5:30**

7% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 6% Internet database
- 4% Publications database
- Crossref database
- Crossref Posted Content database
- 4% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less then 10 words)

7% Overall Similarity

Top sources found in the following databases:

- 6% Internet database
- 4% Publications database
- Crossref database
- Crossref Posted Content database
- 4% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	University of Nottingham on 2018-04-29 Submitted works	<1%
2	journals.sagepub.com Internet	<1%
3	huggingface.co Internet	<1%
4	fastercapital.com Internet	<1%
5	mdpi.com Internet	<1%
6	udspace.udel.edu Internet	<1%
7	Georgia State University on 2023-03-09 Submitted works	<1%
8	The University of Buckingham on 2023-06-21 Submitted works	<1%

9	Zambia Centre for Accountancy Studies on 2020-12-20 Submitted works	<1%
10	apps.dtic.mil Internet	<1%
11	ierj.in Internet	<1%
12	thetechedvocate.org Internet	<1%
13	Chen Chen, Yi Han, Yan Chen, Hung-Quoc Lai, Feng Zhang, Beibei Wan... Crossref	<1%
14	The University of Manchester on 2023-06-14 Submitted works	<1%
15	artemis.cslab.ece.ntua.gr:8080 Internet	<1%
16	yamanashi.repo.nii.ac.jp Internet	<1%
17	University of Sheffield on 2021-09-22 Submitted works	<1%
18	liu.diva-portal.org Internet	<1%
19	research-collection.ethz.ch Internet	<1%
20	epublications.marquette.edu Internet	<1%

21	jwcn-eurasipjournals.springeropen.com Internet	<1%
22	igi-global.com Internet	<1%
23	researchgate.net Internet	<1%
24	Higher Education Commission Pakistan on 2010-08-06 Submitted works	<1%
25	John Irungu, Timothy Oladunni, Andrew C. Grizzle, Max Denis, Marzieh ... Crossref	<1%
26	University of Edinburgh on 2011-08-19 Submitted works	<1%
27	University of Florida on 2016-04-27 Submitted works	<1%
28	core.ac.uk Internet	<1%
29	dspace.cvut.cz Internet	<1%
30	link.springer.com Internet	<1%
31	pubmed.ncbi.nlm.nih.gov Internet	<1%
32	scholar.sun.ac.za Internet	<1%

33	australia24.com.au Internet	<1%
34	electricaltechnology.xyz Internet	<1%
35	D'Onofrio, Anthony, Jr.. "Distributed Learning in a Quantum-Classical H... Publication	<1%
36	Sampaio, Ana Filipa Teixeira. "DL4Malaria: Deep Learning Approaches ... Publication	<1%
37	University of Hertfordshire on 2023-09-18 Submitted works	<1%
38	University of Queensland on 2018-11-05 Submitted works	<1%
39	University of St Andrews on 2024-02-19 Submitted works	<1%
40	dione.lib.unipi.gr Internet	<1%
41	elibrary.tucl.edu.np Internet	<1%
42	interestingengineering.com Internet	<1%
43	open-innovation-projects.org Internet	<1%
44	www2.mdpi.com Internet	<1%

Only_Contect MTech Thesis.pdf

My Files

My Files

Delhi Technological University

Document Details

Submission ID

trn:oid:::27535:59955928

Submission Date

May 25, 2024, 12:00 PM GMT+5:30

Download Date

May 25, 2024, 12:03 PM GMT+5:30

File Name

Only_Contect MTech Thesis.pdf

File Size

2.1 MB

54 Pages

12,023 Words

66,874 Characters

How much of this submission has been generated by AI?

0%

of qualifying text in this submission has been determined to be generated by AI.

Caution: Percentage may not indicate academic misconduct. Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Frequently Asked Questions

What does the percentage mean?

The percentage shown in the AI writing detection indicator and in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was generated by AI.

Our testing has found that there is a higher incidence of false positives when the percentage is less than 20. In order to reduce the likelihood of misinterpretation, the AI indicator will display an asterisk for percentages less than 20 to call attention to the fact that the score is less reliable.

However, the final decision on whether any misconduct has occurred rests with the reviewer/instructor. They should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in greater detail according to their school's policies.

How does Turnitin's indicator address false positives?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be AI-generated will be highlighted blue on the submission text.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

What does 'qualifying text' mean?

Sometimes false positives (incorrectly flagging human-written text as AI-generated), can include lists without a lot of structural variation, text that literally repeats itself, or text that has been paraphrased without developing new ideas. If our indicator shows a higher amount of AI writing in such text, we advise you to take that into consideration when looking at the percentage indicated.

In a longer document with a mix of authentic writing and AI generated text, it can be difficult to exactly determine where the AI writing begins and original writing ends, but our model should give you a reliable guide to start conversations with the submitting student.