IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

NON-VOICE HUMAN SOUND DISEASE DETECTION SYSTEM USING MODIFIED DEEP LEARNING

Gunaprabha K, Student, Ravikumar P, Professor, Department of Biomedical Engineering, Velalar College of Engineering and Technology, Erode, India

Abstract - This innovative project addresses the growing need for efficient healthcare solutions by introducing a Human Sound-Based Diseases Detection System. Leveraging advanced machine learning techniques and signal processing, the system aims to diagnose respiratory conditions like asthma, COVID-19, tuberculosis, and obstructive diseases through the analysis of audio data converted into spectrogram images. The methodology involves utilizing the Librosa library for audio preprocessing, employing the Dense Net architecture for model training, and discussing the selected database's pivotal role. Results indicate the CNN's impressive accuracy improvement from 0.3 to 0.99 during 20 epochs, highlighting the system's ability to make accurate predictions. The discussion emphasizes the need for further validation, consideration of potential biases, and exploration of additional metrics to enhance the system's applicability in non-invasive disease detection

INDEX WORDS: Asthma, Tuberculosis, obstructive, covid-19, Librosa library, CNN.

I. Introduction

The human voice offers subtle cues that can be indicative of a variety of physiological conditions, making it a unique and valuable indicator of an individual's health. The complex relationship between diseases and the human voice is examined in this chapter, with particular attention to the role that vocal traits play in the recognition and comprehension of respiratory disorders. Through exploring this relationship, the Human Sound-Based Diseases Detection System can utilize audio data to support the early identification and tracking of respiratory system-related diseases [4]. In A voice recognition tool is used in voice analytics to record and analyze audio. In addition to translating speech to text, voice analytics software decodes audio signals to understand the speaker's intentions and feelings. A plethora of research exists. There is an abundance of research showing that a person's speech can be affected by multiple physical and mental health conditions. Also, during the speaking, there are 6300+ parameters which become active from which some sets of the parameters are affected in each health condition. They could also make your voice creak or jitter so briefly that it’s not detectable to the human ear. For example, speaking in a more nasal accent, elongating noises, slurring words or even noises that are not audible to the human ear. So in the proposed system we develop a machine learning architecture to classify the diseases based on human sounds.

Voice analytics uses a voice recognition tool to analyze and record an audio. Voice analytics program not only converts speech to text, but can also recognize the sentiments and intent of the speaker by interpreting audio signals. There is an abundance of research showing that a person's speech can be affected by multiple physical and mental health conditions. Also, during the speaking, there are 6300+ parameters which become active from which some set of the parameters are affected in each health condition. They could also make your voice creak or jitter so briefly that it’s not detectable to the human ear. For example, speaking in a more nasal accent, elongating noises, slurring words or even noises that are not audible to the human ear. The below are the advantages of voice analytics, which will help us in detection and prevention of diseases: Point of Care / Rapid Screening: It helps in achieving accurate real time results within minutes (rapidly)rather than hours and ensures that the patients receive the most effective and efficient treatment when and where it is needed with ease to use. Early Warning: It also helps in providing early warning to the users if there is a case of future emergency so that everyone should be safe and protected from the diseases. Disease Surveillance: It is an information which performs the gathering, evaluating, and interpretation of large data from a variety of sources. Preventive Care and Wellness: It helps in detecting or preventing serious diseases before they become crucial. This can lead to save us from future problems. From our Voice production system to Throat, there are a total of 18 articulation points. Each point has distinct features.

Whenever a healthy person communicates, the individual's speech characteristics are relatively usual. But when a person suffers from a particular health condition, certain parameters of the speaking voice of the spoken phrases impair them. The careful analytics of the speaking voice has a potential to map the underlying disease conditions. Voice analytics is used by the, speech pathologists to identify voice conditions by auditory perceptual criteria such as breathability, gruffness and harshness. [11] These diagnoses, on the other hand, depend on the knowledge of clinicians and require subjective recognition. It's easy to mix up pathologies with distinct symptoms with those that are often referred to as hoarse. Also, when new technologies are used, for example, the same problem occurs when such speakers display a reflex motion in their supra-glottal cavity, resulting in incorrect judgments. The problem remains in the absence of a doctor-to-patient recommendation system that alerts the doctor to the patient's condition and provides excellent care to the patient, since they have access to a doctor at all times. The problematic statement is building a model for disease prediction using voice recognition. Since the lack of a normalized text for sample collection is a significant issue, our project aims to make it language independent.

II. Types of disease

Respiratory conditions encompass a wide spectrum of disorders, each presenting with distinct symptoms and effects on vocal patterns. This section provides an overview of various respiratory diseases, emphasizing their impact on the human voice. From chronic conditions like asthma to infectious diseases like COVID-19 and tuberculosis, the diversity of these disorders necessitates a nuanced approach in their identification and characterization.[2]

A. VOICE CHARACTERIZATION

In order to develop an effective diagnostic system, a detailed examination of each respiratory disease and its associated voice characteristics is essential. This sub-section delves into the specifics of diseases such as asthma, COVID-19, tuberculosis, and obstructive diseases. Understanding how each condition manifests in vocal patterns is crucial for training the Human Sound-Based Diseases Detection System to accurately identify and classify these diseases based on audio data.[12]

1. ASTHMA

Common triggers for asthma include: Indoor allergens, such as dust mites, mold, and pet dander or fur. Outdoor allergens, such as pollens and mold. Emotional stress, such as intense anger, crying, or laughing

FEATURES

Noise ratio	-1.58E05
spectral contrast means	18.83571
mfcc mean	5.65379
closed quotient	0.99725
open quotient	0.00275
Label	0

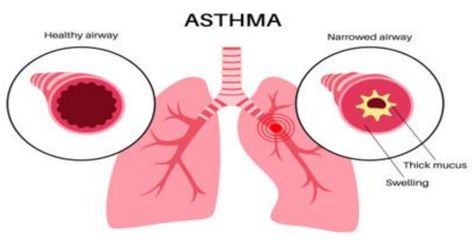

Fig2.1 Asthma

2. COVID

The virus spreads by respiratory droplets released when someone with the virus coughs, sneezes, breathes, sings or talks. more of the air sacs can become filled with fluid leaking from the tiny blood vessels in the lungs.

FEATURES

Noise ratio	8.61E-06
spectral contrast mean	17.89167
mfcc mean	-10.7872
closed quotient	0.999813
open quotient	0.000188
label	1

FIG 2.2 Covid

3. TUBERCULOSIS:

Tuberculosis (TB) is caused by bacteria. It can spread through close contact with people who have TB and have symptoms (active TB). When someone with active TB coughs, they release small droplets containing the bacteria.

FEATURES:

Noise ratio	-4.13E-06
spectral contrast mean	15.87154
mfcc mean	-21.1646
closed quotient	0.99925
open quotient	0.00075
label	2

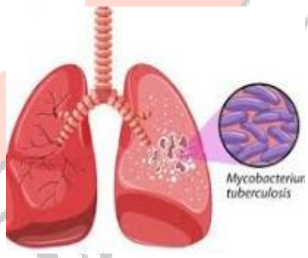

Fig 2.3 Tuberculosis

4. OBSTRUCTIVE

Obstructive uropathy can occur due to a variety of factors. Compression can lead to damage to your kidneys and ureters. Temporary or permanent blockages in your ureter or urethra, through which urine exits your body, can result from: injuries such as a pelvic fracture.

FEATURES

Noise ratio	3.71E-06
spectral contrast mean	14.21938511
mfcc mean	-35.99
closed quotient	0.999938
open quotient	6.25E-05
label	4

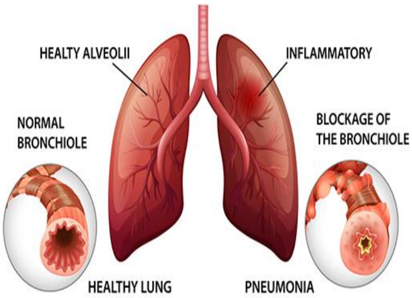

Fig 2.4 Obstructive

III. METHODS

A. SOFTWARE USED

The methodology section is a critical component that outlines the tools, frameworks, and procedures employed in the development of the Human Sound-Based Diseases Detection System [6]. This sub-chapter begins by providing a detailed examination of the software components selected for various stages of the project.

The choice of the Librosa library for audio data preprocessing is rooted in its versatility and efficiency in handling audio data in Python. Librosa facilitates the transformation of raw audio samples into spectrogram images, capturing the temporal and frequency characteristics essential for disease classification. Furthermore, the decision to implement the deep learning model using a framework like TensorFlow or Pitch stems from their robust capabilities in building and training complex neural network architectures. The selection of TensorFlow or PyTorch is driven by factors such as ease of use, community support, and compatibility with the chosen neural network architecture – Dense Net.[13]

B. FLOW CHART

A comprehensive understanding of the project's workflow is crucial for ensuring transparency and reproducibility. This section introduces a detailed flowchart that delineates each block of the Human Sound-Based Diseases Detection System. The flowchart begins with the data acquisition block, involving the collection of diverse audio datasets representing different respiratory conditions.

The subsequent preprocessing block utilizes the Librosa library to convert raw audio data into spectrogram images, preserving essential information for disease characterization.

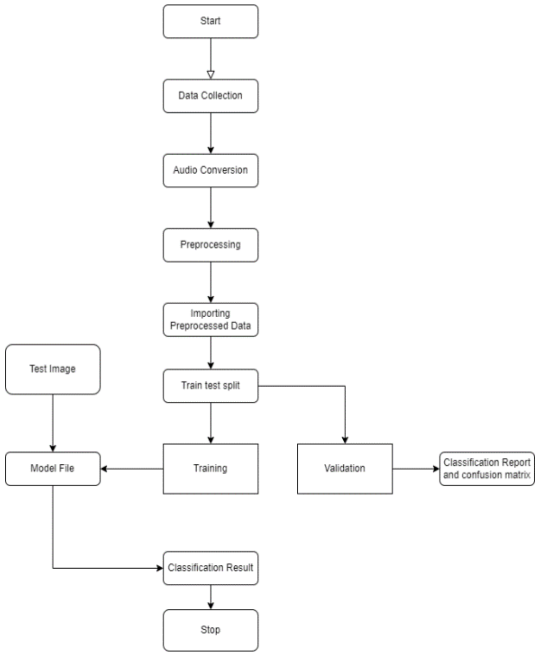

Fig 3.1 Flow Chart

The core of the system lies in the model training block, where the Dense Net architecture is employed to learn intricate patterns and features from the spectrogram images. The choice of Dense Net is rationalized by its ability to capture complex relationships in images, ensuring the model's proficiency in recognizing distinct spectrogram patterns associated with each respiratory condition. The classification block utilizes the trained model to classify new audio samples into specific disease categories, providing a practical and efficient means of diagnosis. Each block's role in the overall functionality is thoroughly explained, emphasizing the iterative and interconnected nature of the process.[14]

C. DATABASE USED

The selection of an appropriate database is a pivotal decision in training and evaluating the model's performance. This sub-section discusses the considerations and criteria for choosing the database, including factors such as diversity, size, and relevance to the targeted respiratory conditions. The database's role in shaping the model's ability to generalize and accurately classify diseases is highlighted, emphasizing the importance of representative and well-curated datasets.

D. ARCHITECTURE USED

The architectural design of the Human Sound-Based Diseases Detection System is a key determinant of its effectiveness. This section provides an in-depth explanation of the chosen architecture, emphasizing the Dense Net model's suitability for the task at hand. Dense Net's unique connectivity pattern and feature reuse properties are discussed, showcasing how these characteristics contribute to the model's ability to capture intricate patterns in spectrogram images.

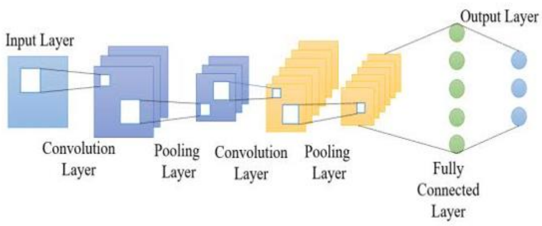

Fig 3.2 Dense Net architecture

By offering detailed insights into the methodology, this chapter ensures clarity in the development process of the Human Sound-Based Diseases Detection System. The synergy between software tools, workflow blocks, database selection, and the chosen neural network architecture is crucial for the successful implementation of an accurate and robust disease detection system based on audio data analysis.[21]

IV. RESULTS & DISCUSSION

A. Audio to spectrogram image converted image:

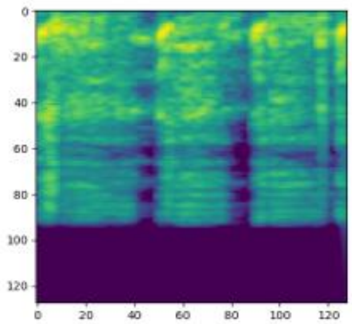

Fig4.1 Asthma audio signal in spectrogram representation of image format

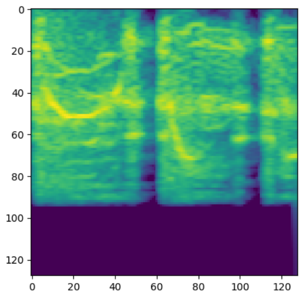

Fig4.2 Covid audio signal in spectrogram representation of image format

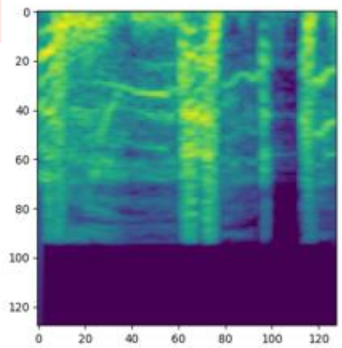

Fig4.3 Obstructive audio signal in spectrogram representation of image format

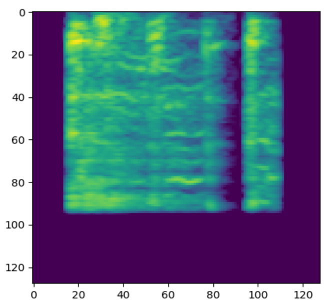

Fig4.4 Tuberculosis audio signal in spectrogram representation of image format

B. VARIOUS TYPES OF OUTPUT

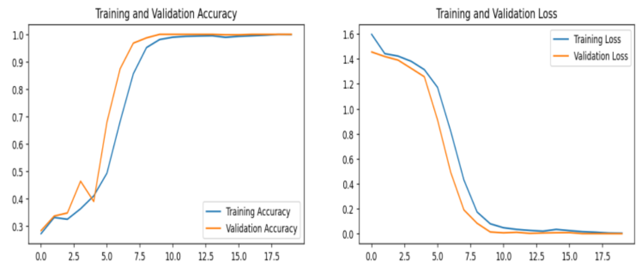

Fig 4.5 Accuracy and Loss in graphical representation for Training and Validation of CNN Model

The culmination of the Human Sound-Based Diseases Detection System is reflected in the outcomes obtained during the training of the Convolutional Neural Network (CNN) over 20 epochs. This section meticulously details the various types of outputs, primarily focusing on accuracy and loss metrics, shedding light on the system's performance evolution.

Initially, the model exhibited a modest performance with an accuracy of 0.3 and a loss of 1.6. This early stage indicates the model's initial struggle to capture the intricate patterns in the training data. However, as the training progressed, a notable improvement unfolded. The accuracy surged impressively to 0.99, indicating that the CNN learned to make highly accurate predictions on the given dataset. Simultaneously, the loss decreased substantially to 0.01, signifying those the model's predictions aligned more closely with the actual ground truth values.

These metrics suggest that the CNN effectively learned and adapted its internal representations to the patterns present in the training data. The accuracy reaching 0.99 indicates that the model can make correct predictions for a significant majority of the samples in the training set. The low loss value of 0.01 further corroborates the model's proficiency in minimizing the disparity between predicted and actual values.

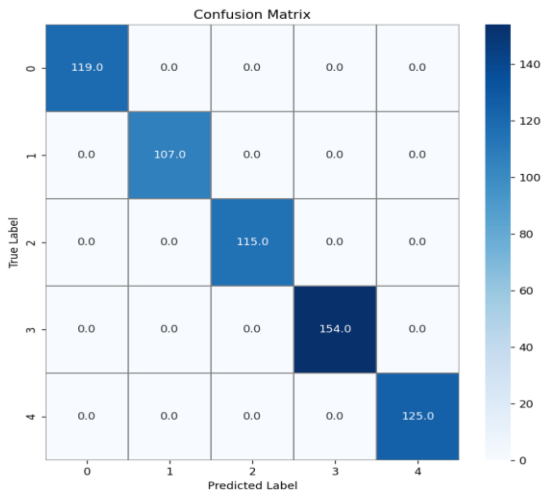

Fig 4.6 Confusion matrix

The acquired confusion matrix is result for testing the model file created by CNN architecture using the give datasets. This confusion matrix explains in the 1st row the data file in the 0th labeled document is taken and finds 119.0 images as asthma out of 119.0 images with 100% accuracy and in the 2nd row the data file in the 1st labeled document is taken and finds 107.0 images as covid out of 107.0 images with 100% accuracy and in the 3rd row the data file in the 2nd labeled document is taken and finds 115.0 images as tuberculosis out of 115.0 images with 100% accuracy and in the 4th row the data file in the 3rd label document is taken and finds 154.0 images as asthma out of 154.0 images with 100% accuracy and in the 5th row the data file in the 4th label document is taken and finds 125.0 images as asthma out of 125.0 images with 100% accuracy.

Sample Input	Method	Predicted Disease	Total Accuracy	Previous Work Method	Predicted Diseases	Result & Accuracy
Audio 1	CNN	Tuberculosis	99.9%	CNN	LUNG DISEASES	99.76 %
Audio 2		Asthma		MFCC		99.79 %
Audio 3		Covid		ANN		98.2
Audio 4		Healthy		VGG		69%
Audio 5		Obstructive		SVM		40.9%

Table4.1 Result Comparison

C. DISCUSSION OF OBTAINED RESULTS

The positive trend observed in accuracy and loss metrics during training signifies a well-trained CNN, showcasing its capability to generalize and make accurate predictions on new, unseen data. However, it is crucial to extend the evaluation beyond the training set to ensure the model's generalizability. This necessitates validation on a separate validation set or test set, offering a more comprehensive assessment of the system's effectiveness.

Additionally, the discussion delves into considerations regarding potential overfitting, a phenomenon where the model becomes overly specialized to the training data and struggles with new, unseen samples. Monitoring for signs of overfitting during training and employing regularization techniques are essential to enhance the model's robustness.

The obtained results prompt considerations for further improvements and optimizations. Exploring additional performance metrics, such as precision, recall, and F1 score, can provide a more nuanced evaluation of the model's effectiveness in disease classification. Furthermore, addressing potential biases in the dataset and ensuring diversity in the training data are crucial steps in enhancing the system's applicability across different demographic groups. the positive outcomes observed during training indicate the promising potential of the Human Sound-Based Diseases Detection System. The discussion provides a critical analysis of the achieved results, paving the way for future enhancements, validations, and real-world applications of the proposed system in non-invasive and efficient disease detection.

V. CONCLUSION

In conclusion, the innovative system for early detection of respiratory diseases through voice analytics and deep learning presents a promising solution to the challenges associated with traditional diagnostic methods. by harnessing the power of voice data, this system offers early detection, comprehensive analysis, cost-effectiveness, user-friendliness, and the potential for remote monitoring. the integration of over 6300 parameters during analysis ensures a nuanced understanding of an individual's health, marking a significant advancement in the field of respiratory disease diagnostics.

A. FUTURE SCOPE

The future scope of this project extends to further refinement and expansion of the system's capabilities. continuous improvement in machine learning algorithms and voice recognition technologies can enhance the accuracy and reliability of disease detection. collaboration with healthcare professionals and integration with electronic health records could further streamline the diagnostic process. additionally, the system could be adapted to detect other health conditions beyond respiratory diseases, broadening its applicability and impact on preventive healthcare. further research and development in this direction hold the potential to revolutionize early disease detection and healthcare accessibility on a global scale.

Reference

[1] Wen-Hung Liao, Yu-Kai Lin," Classification of non-speech human sounds: Feature selection and snoring sound analysis",2009 IEEE International Conference on Systems Man and Cybernetics

[2] Trevor R. Agus, Simon J. Thorpe, Clara Suied, Daniel Pressnitzer, "Characteristics of human voice processing",2010 IEEE International Symposium on Circuits and Systems (ISCAS)

[3] Moe Yaegashi, Tetsuji Ogawa, "Do You Know How Humans Sound? Exploring a Qualification Test Design for Crowdsourced Evaluation of Voice Synthesis Quality",2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)

[4] Yuan Gong, Jin Yu, James Glass, "Vocal sound: A Dataset for Improving Human Vocal Sounds Recognition", ICASSP 2022 - 2022 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)

[5] Sung-Eun Kim, Tae Wook Kang, Junghwan Hwang, Sungweon Kang, Kyunghwan Park, " Sound transmission through the human body with digital weaver modulation (DWM) method",2014 IEEE International Systems Conference Proceedings

[6] Lacrimioara Grama, Lorena Muscar,Corneliu Rusu,"Sound Classification Algorithms for Indoor Human Activities",2021 16th International Conference on Engineering of Modern Electric Systems (EMES)

[7] Daria Soboleva, Ondrej Skopek, Me1rius ajgaledk, Victor rune, Felix Weissenberger, Julia Proskurnia, Bogdan Prisacari, Daniel Valcarce, Justin Lu, Rohit Prabhavalkar, Balint Miklos, "Replacing Human Audio with Synthetic Audio for on-Device Unspoken Punctuation Prediction", ICASSP 2021 - 2021 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)

[8] Nathaniel T. Greene, Gary D. Paige, "Influence of sound source width on human sound localization",2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society

[9] Ji-Yong Lee, Doik Kim, Bum-Jae You, Nakju Lett Doh, "Human localization based on the fusion of vision and sound system",2011 8th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)

[10] J.D. Hoyt, H. Wechsler, "Detection of human speech in structured noise", Proceedings of ICASSP 94. IEEE International Conference on Acoustics Speech and Signal Processing

[11] Georgios Petmezas, Grigorios-Aris Cheimariotis, Leandros Stefanopoulos, Bruno Rocha, Rui Pedro Paiva, Aggelos K. Katsaggelos, and Nicos Maglaveras, " Automated Lung Sound Classification Using a Hybrid CNN-LSTM Network and Focal Loss Function", 2022 Sensors in Physiological Sound Acquisition and Processing Volume II

[12] Kranthi Kumar Lella, Alphonse Pja "Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: Cough, voice, and breath" 2022 Alexandria Engineering Journal.

[13] Dong-Min Huang, Jia Huang, Kun Qiao, Nan-Shan Zhong, Hong-Zhou Lu & Wen-Jin Wang "Deep learning-based lung sound analysis for intelligent stethoscope" 2023. Sensors in Physiological Sound Acquisition and Processing Volume.

[14] Abdullah M. Almuhaideb "Cough Sound Detection and Diagnosis Using Artificial Intelligence Techniques Challenges and Opportunities" 2023. (APSIPA ASC)

[15] Tong Xia, Jing Han "Exploring machine learning for audio-based respiratory Condition screening: A concise review of databases, methods, and open issues" 2022. IEEE International

[16] John H. L. Hansen, Mahesh Kumar Nandwana and Navid Shokouhi "Analysis of human scream and its impact on text-independent speaker verification" 2018 The Journal of the Acoustical Society of America

[17] Rahna.K.M1 and Baby.C.J "Detection of Non-Speech Human Sounds for Surveillance "2018 International Journal of Computer Science and Engineering Communications

[18] burak uzkent1 , buket d. barkana1,∗ and hakan cevikalp2" non-speech environmental sound classification using svms with a new set of features" international journal of innovative computing, 2021 information and control icic international

[19] Riikka Mo¨tto¨ nen,a,b,* Gemma A. Calvert,c Iiro P. Ja¨a¨skela¨inen,a,b,d Paul M. Matthews,e Thomas Thesen,c Jyrki Tuomainen,a,f and Mikko Sams a, "Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus "2021 neuro image

[20] Shanq-Jang Ruan 1 and Lieber Po-Hung Li "Environmental Noise Classification with Inception-Dense Blocks for Hearing Aids" 2021 sensors

[21] Manasi Bendale1 , Jaya Soni2 , Garvit Chhugani3 , Nikhil Sahu4 , Nitin Choubey5 , Santosh Bothe6 "Voice Based Disease Identification System" 2018 Turkish Journal of Computer and Mathematics Education

[22] Muntasir Hoq 1 , Mohammed Nazim Uddin 1,* and Seung-Bo Park 2"Vocal Feature Extraction-Based Artificial Intelligent Model for Parkinson's Disease Detection" 2023, diagnostics mdpi

[23] Alyaa Hamel Sfayyih1 , Nasri Sulaiman1* and Ahmad H. Sabry2" A review on lung disease recognition by acoustic signal analysis with deep learning networks" 2023 Department of Electrical and Electronic Engineering,

[24] Nihad Karim Chowdhury a,* , Muhammad Ashad Kabir b , Md. Muhtadir Rahman a , Sheikh Mohammed Shariful Islam c "Machine learning for detecting COVID-19 from cough sounds": 2022 An ensemble-based MCDM method computers in biology medicine

[25] Gabriela Augustinov, Patrick Fischer, Volker Gross, Ulrich Koehler, Keywan Sohrabi, and Seyed Amir Hossein Tabatabaei* "Automatic Detection and Classification of Cough Events Based on Deep Learning" 2020 biomedical Current Directions in Biomedical Engineering