Syrian Arab Republic
Ministry of Higher Education
and Scientific Reserch
Syrian Virtual University

الجمهورية العربية السورية
وزارة التعليم العالي والبحث العلمي
الجامعة الافتراضية السورية

Respiratory Diseases Detection and Classification Based on Respiratory Voice Using Artificial Intelligence Methods

(A thesis submitted as a fulfilment of requirements for a Master's degree in Bioinformatics)

By:

Razan Dkhan

Supervised By:

Dr. Raouf Hamdan

F23-2024

Table of Contents:

Contents

Table of Abbreviation

Abbreviation	Meaning
RS	Respiratory System
COPD	Chronic obstructive pulmonary disease
DL	Deep Learning
ML	Machine Learning
RD	Respiratory Diseases
ILD	Interstitial lung disease
OSA	Obstructive sleep apnea
ANI	Artificial Narrow Intelligence
AGI	Artificial General Intelligence
ASI	Artificial Super Intelligence
ANN	Artificial Neural Networks
RNN	Recurrent Neural Networks
CNN	Conventional Neural Networks
LSTM	Long Short-Term Memory
BERT	Bidirectional Encoder Representations Transformer
CLASS	Computerized Lung Auscultation – Sound System
MFCC	Mel-Frequency Cepstral Coefficients
FFT	fast forier transformation
DCT	Discrete cosine transform
Tonnetz	Tonal Centroid Features
GPT	generative pre-trained transformers

Table of Figures

Table of Tables

Abstract

Background: Respiratory illnesses are a significant global health issue, accounting for a large proportion of smokers, addiction, air pollution, increased CO2, occupational hazards like chemical fumes, and common allergens. Despite advancements in detection and treatment, respiratory diseases remain a leading cause of lung cancer and deaths globally.

Deep learning and transformers, branches of artificial intelligence, have emerged as valuable tools for predicting respiratory diseases (RD) including COPD, asthma, and URTI, for early detection and treatment. Harnessing the power of DL and transformers in respiratory issues prediction holds promise for advancing personalized medicine and optimizing treatment strategies.

Aim of Study: The aim of this study is to find high accuracy and sensitive algorithms capable of predicting RD based on respiratory sounds. The study considers respiratory sounds to reduce harmful and surgical diagnosis for quickly intervention in the patient's treatment protocol for less mortality cases.

Material and Methods: This study utilized the respiratory sound database; The data is an open-source dataset that was compiled by two collaborative research teams based in Portugal and Greece. It includes a total of 920 recordings obtained from 126 subjects, up to 6898 respiratory cycles. The data contains patient ID, age, sex, diagnosis, acquisition tool, and the chest location.

The features were extracted and the dataset was split into training and testing sets for model development and evaluation (DL and transformer). Data preprocessing techniques were applied and Python libraries facilitated data manipulation and analysis.

Results: The accuracy of the algorithms varied, RNN model achieved the lowest accuracy of 87.68%, while transformer achieved better results up to 90% for MFCC feature. The best accuracy shown by CNN model which reached to 97.5%. The findings highlight the superior performance of CNN and the lowest performance of RNN model as the CNN is the most common model used for audio signal.

Conclusion: The study underscores the importance of selecting appropriate classification algorithms for predicting RD. The CNN model and Transformer offer promising results, while RNN may not be the most effective choice. These findings contribute to the field of respiratory system prognosis and provide insights for improving personalized treatment strategies.

Thermotical Review

Chapter 1- Respiratory System

1.1 Introduction:

The respiratory system is the organs and other parts of our body responsible for breathing operation, when we exchange oxygen and carbon dioxide.

All the cells in our body need oxygen to work. As they take in oxygen, they release carbon dioxide, which is called a "waste gas". It goes into the bloodstream and gets carried to our lungs, which will be exhaled while we breathe out.

This vital function is called "gas exchange," and our body is set up to do it automatically. [44]

Problems with our respiratory system can reduce the oxygen that reaches our lungs, thus not all cells in our body will have oxygen which can make us unwell.

1.2 Parts of respiratory system:

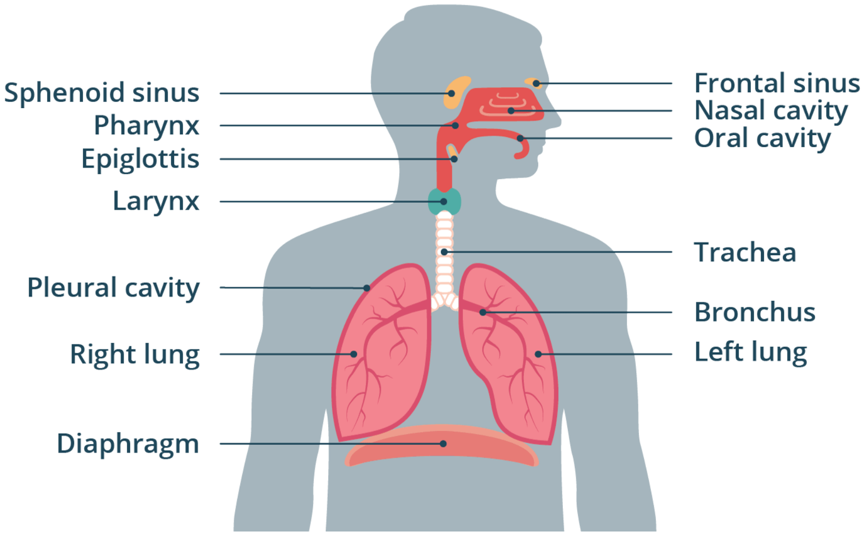


Figure 1.1: The respiratory system parts [19]

The respiratory system is divided into two parts, upper and lower.

The upper respiratory tract is made up of:

Nose and nasal cavity, Sinuses, Throat (pharynx), Mouth, Voice box (larynx).

The parts of lower respiratory tract are:

Windpipe (trachea), Diaphragm, Lungs, Bronchial tubes/bronchi, Bronchioles, Air sacs (alveoli), and Capillaries

To define them:

- ✓ The **NASAL CAVITY** (nose) is the main entrance of outside air into respiratory system. The hairs that line the inside wall are part of the air-cleansing system.
- ✓ Air can also enter through the **ORAL CAVITY** (mouth), especially if there is a mouth-breathing habit or the nasal passages may be temporarily blocked.
- ✓ **Lungs** remove the oxygen and pass it through bloodstream, where it's carried off to the tissues and organs that allow human to walk, talk, and move.

  Lungs also take carbon dioxide from blood and release it into the air when breathing out.
- ✓ The **SINUSES** are hollow spaces in the bones of the head. Small openings connect them to the nasal cavity. The sinuses help to regulate the temperature and humidity of air we breathe in, as well as to lighten the bone structure of the head and to give tone to the voice.
- ✓ The **PHARYNX** (throat) collects incoming air from the nose and passes it downward to trachea (windpipe).
- ✓ The **LARYNX** (voice box) is a hollow tube that's about 4 to 5 centimeters (cm) in length and width. It contains vocal cords and lets air pass from throat (Pharynx) to trachea on the way to lungs. When moving air breathed in and out, Larynx creates voice sounds so it's called voice box.
- ✓ The **TRACHEA** (windpipe) is the passage leading from pharynx to the lungs.
- ✓ The **DIAPHRAGM** is the strong wall of muscle that separates chest cavity from the abdominal cavity. By moving downward, it creates suction to draw in air and expand the lungs.
- ✓ The smallest section of the bronchi is called **BRONCHIOLES**, at the end of which are the alveoli (plural of alveolus). [32]

1.3 Functions of respiratory system:

The main function of the respiratory system is to pull in oxygen for body's cells and get rid of carbon dioxide, a waste product.

The other functions:

- Warming the air breathing in to match the body temperature, and moisturizes the air to bring it to the humidity level the body needs.
- Protects the body from particles breathing in as it can block harmful germs and irritants from getting in, or push them out if they do get in.
- Allows to talk. Because air vibrates vocal cords which makes sounds.
- Helps to smell. Breathing in air moves its molecules past the olfactory nerve, which sends messages to the brain about the way something smells.
- Balances level of acidity in the body. Too much carbon dioxide lowers blood's pH, making it acidic. By removing carbon dioxide, the respiratory system helps maintain the acid-base balance in the body. [18]

1.4 The Mechanism of Respiratory System:

breathing in occurs by contracting the diaphragm, which is a flat muscle at the base of the chest. This causes the chest to expand, drawing air in.

breathing air in and out through the nose and mouth and the air is warmed and moistened along the way.

The air passes through the larynx, which contains the vocal cords that allow to talk.

Air then passes into through the upper airways, including the trachea (windpipe) and bronchi to reach lungs. The lining of the respiratory tract makes mucus to trap foreign particles.

In lungs, air sits in small air sacs called alveoli, which are right next to blood vessels Oxygen from the air breathed in travels from alveoli into the bloodstream. Carbon dioxide travels the other way, from bloodstream into alveoli. Then the carbon dioxide goes out through breathing out. [30]

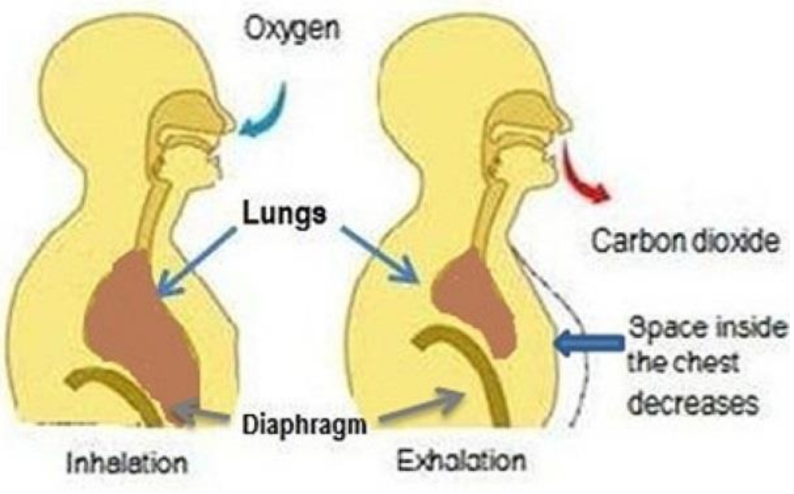


Figure 1.2: The workflow of respiratory system. [30]

1.5 The Respiratory Sounds:

The Respiratory Sounds also known as lung sounds or breath sounds, are the specific sounds generated by the movement of air through the respiratory system. These may be easily audible or identified through auscultation of the respiratory system through the lung field with a stethoscope as well as from the spectral characteristics of lung sounds.

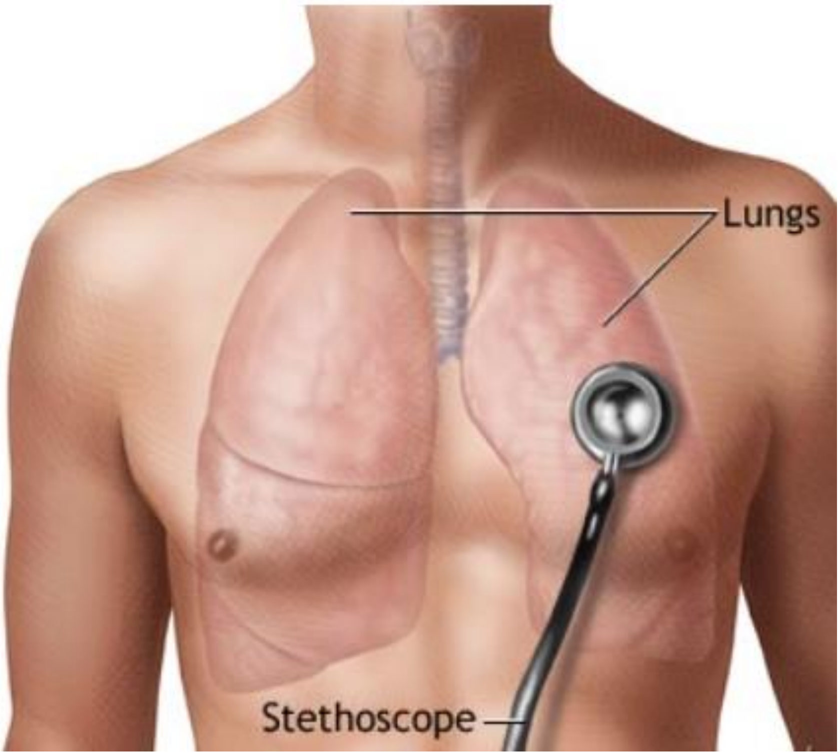


Figure 1.3: Respiratory sound acquisition [16]

Using a stethoscope, the health care provider may hear normal breathing sounds, decreased or absent breath sounds, and abnormal breath sounds:

1.5.1 Normal respiratory sounds:

1) **Normal lung or "vesicular" sounds** are soft, nonmusical, and audible over almost entire peripheral lung zones during inspiration and early expiration. They are produced by turbulent air flow through lobar and segmental bronchi. In disease states, there may be diminished intensity due to decreased generation of sound energy, impaired sound transmission, or both. Decrease in sound generation may be due to impaired respiratory drive or impaired flow of air to the peripheral airways (foreign body or obstructive airway diseases). Impaired transmission of sounds may be due to the presence of fluid or air in the pleural space, consolidated lung, large bullae in patients with emphysema, or by chest wall deformities or obesity.

2) **Normal tracheal sounds** are hollow nonmusical sounds with a wide spectrum of frequencies that are clearly heard at the suprasternal notch or the lateral neck in both respiratory cycles. **Pathologic tracheal or "bronchial" sounds** are audible over peripheral lung areas and may suggest lung consolidation (due to inflammation, infection, hemorrhage, protein, or malignancy). In patients with upper airway obstruction, tracheal sounds may become musical and can present as either stridor or localized wheeze.[1]

1.5.2 Absent or decreased sounds:

Mean:

- Air or fluid in or around the lungs (such as pneumonia, heart failure, and pleural effusion)
- Increased thickness of the chest wall
- Over-inflation of a part of the lungs (emphysema can be a reason)
- Reduced airflow to part of the lungs

1.5.3 Abnormal breath sounds:

Types of abnormal breath sounds:

1) **Crackles** are nonmusical, short (<0.25 seconds), explosive respiratory sounds heard mostly during inspiration, caused by the sudden equalization of gas pressures between two areas of the lung. They occur during the

opening of previously closed small airways. Crackles may be transiently apparent in healthy people but disappear after a few deep inspirations.

a) **Fine crackles,** formerly termed crepitations or “velcro rales,” are heard mid-to-late inspiration mostly in dependent lung regions, uninfluenced by cough or body position, and not transmitted to the mouth. These high-pitched sounds may be due to pulmonary fibrosis, congestive heart failure, or pneumonia. Of note, fine crackles are minimal or absent in sarcoidosis, as the disease affects mostly central lung zones.[1]

b) **Coarse crackles** are heard early in inspiration and throughout expiration, can be transmitted to the mouth, and can change with cough, but they are not influenced by changes in body position. These low-pitched sounds are commonly observed in the setting of bronchiectasis and other conditions characterized by secretions in the airways.

2) **Wheezes** and **rhonchi** are musical continuous breath sounds (>0.25 seconds), which may be high-pitched (wheezes) or low-pitched (rhonchi) and are generally audible during expiration. Wheezes (hissing, whistling sounds) are produced by the turbulent flow of air through narrowed airways, while rhonchi are mainly caused by secretions present in the airways. **Expiratory wheeze** is mostly caused by narrowing of the airways within the chest, which can occur in the setting of asthma, chronic obstructive pulmonary disease, aspiration of gastric contents, or heart failure. Of note, localized wheeze may be due to a focal process, including a tumor, foreign body, or mucous plug.

3) **Stridor** is a particularly loud, high-pitched, continuous sound, more clearly heard on inspiration over the upper airways or sometimes even without a stethoscope. This sound is caused by large airway narrowing and may indicate obstruction of the larynx or trachea. Stridor may be heard in patients with vocal cord dysfunction, epiglottitis, airway edema, anaphylaxis, laryngotracheitis, extrinsic compression of the trachea, or a foreign body.

4) **Squawk, also known as “squeak,”** is a mixed sound consisting of short wheezes accompanied by crackles that are heard in the middle to the end of inspiration. Squawks are most frequently present in patients with hypersensitivity pneumonitis and less often in patients with other interstitial lung diseases, bronchiectasis, or pneumonia.[1]

5) **Pleural friction rub** is caused by the rubbing of the parietal and visceral layers of the pleura due to the deposition of fibrin in the course of an inflammatory or neoplastic process. This is generally biphasic in nature and heard best in basal and axillary regions.

6) **Rales. Small clicking, bubbling**, or **rattling sounds** in the lungs. They are heard when a person breathes in (inhales). They are believed to occur when air opens closed air spaces. Rales can be further described as moist, dry, fine, and coarse. [1]

1.6 Vocal Resonance

Normal lung tissue acts as a low-pass filter in that it allows low-frequency sounds to move through easily while filtering high-frequency sounds. Pathological lung tissue can transmit higher frequency sounds more efficiently; this occurs when a normally air-filled lung becomes occupied by another material, such as fluid. Physicians can exploit this phenomenon through the physical exam. Tests used to detect this phenomenon, known as vocal resonance, include bronchophony, egophony, and whispered pectoriloquy. To test for these, the clinician places their stethoscope over symmetric areas of the patient's chest and asks the patient to speak. The clinician usually would hear an unintelligible, distant, and muffled vocal sound. In bronchophony, the voice appears closer and louder. Egophony occurs when pathological lung tissue distorts vowel sounds and makes them more nasal in quality, and therefore makes the sound of a hard E heard as an A, referred to as "E to A changes." Pectoriloquy describes the finding of a clear and intelligible sound when the patient whispers; it usually is unclear and unintelligible. [43]

1.7 Respiratory Disorders:

Respiratory sounds are important indicators of respiratory health and respiratory disorders. The sound emitted when a person breathes is directly related to air movement, changes within lung tissue and the position of secretions within the lung.

Here are some respiratory disorders and its causes:

1- Asthma:

Asthma is a chronic disease that causes the airways of the lungs to swell and narrow. It leads to breathing difficulty such as wheezing, shortness of breath, chest tightness, and coughing.

Asthma is caused by swelling (inflammation) in the airways. An asthma attack occurs when the lining of the air passages has become swollen and the muscles surrounding the airways become tight. This narrowing reduces the amount of air that can pass through the airway.

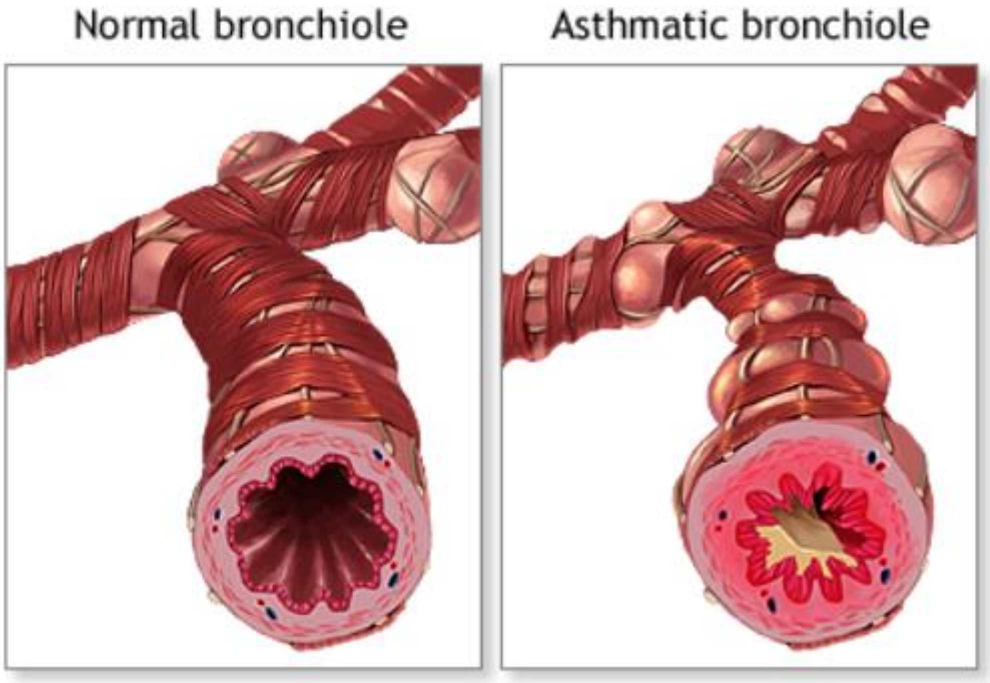


Figure 1.4: Normal and Asthmatic bronchiole [15]

2- Chronic obstructive pulmonary disease (COPD): is a common lung disease which makes it hard to breathe.

There are two main forms of COPD:

- Chronic bronchitis, which involves a long-term cough with mucus
- Emphysema, which involves damage to the lungs over time

The main reason for COPD is smoking which is the most issues that will give you COPD, and the risk factors for COPD are:

Exposure to certain gases or fumes in the workplace.

Exposure to heavy amounts of secondhand smoke and pollution.

Frequent use of a cooking fire without proper ventilation.

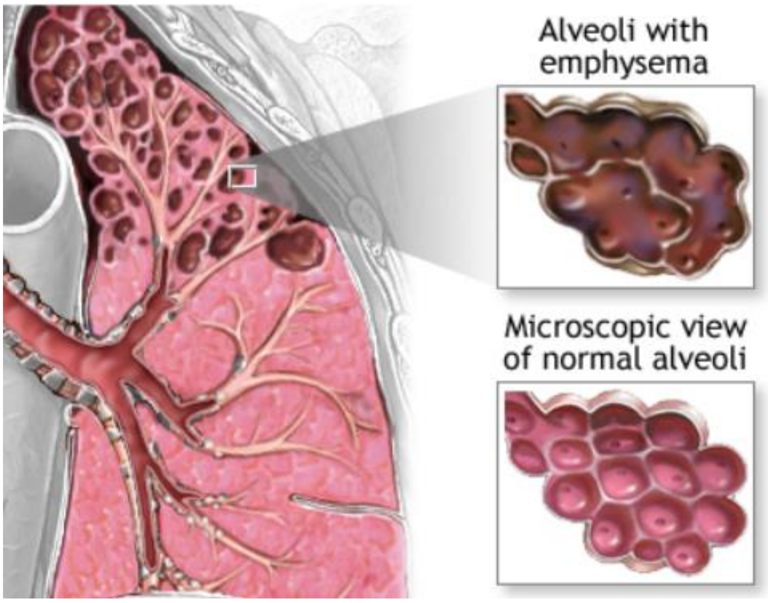


Figure 1.5: The difference between emphysema and normal alveoli [13]

3- Acute bronchitis: is swelling and inflamed tissue in the main passages that carry air to the lungs (bronchi) which narrows the airway to make it harder to breathe. Other symptoms of bronchitis are a cough and coughing up mucus. Acute means the symptoms have been present only for a short time.

The main reason for Acute bronchitis is the cold or flu-like illness. The bronchitis infection is usually caused by a virus. At first, it affects the nose, sinuses, and throat. Then it spreads to the airways that lead to your lungs.[48]

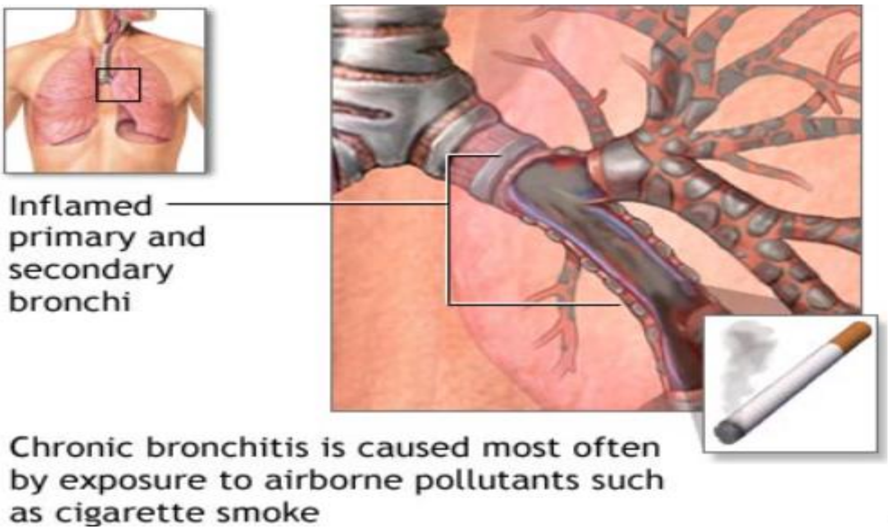

Inflamed primary and secondary bronchi
Chronic bronchitis is caused most often by exposure to airborne pollutants such as cigarette smoke

Figure 1.6: The effect of smoking on primary and secondary bronchi [14]

4- Bronchiectasis :is a disease in which the large airways in the lungs are damaged. This causes the airways to become permanently wider.

Bronchiectasis can be present at birth or infancy or develop later in life.

The main reason for Bronchiectasis is inflammation or infection of the airways that keeps coming back. The other causes of it:

- Allergic lung diseases
- Leukemia and related cancers
- Immune deficiency syndromes
- Primary ciliary dyskinesia (another congenital disease)
- Infection with non-tuberculous mycobacteria
- A complication of bronchiolitis obliterans
- Asthma or chronic obstructive lung disease (uncommon) [13]

5- Interstitial lung disease (ILD): is a group of lung disorders in which the lung tissues become inflamed and then damaged.

The main reason for interstitial ling diseases is that lungs contain tiny air sacs (alveoli), which is where oxygen is absorbed. These air sacs expand with each breath.
Cigarette smoking may increase the risk of developing some forms of ILD and may cause the disease to be more severe.

6- Pneumonia is a breathing (respiratory) condition in which there is an infection of the lung.
Pneumonia is a common illness that affects millions of people each year in the United States. Germs called bacteria, viruses, and fungi may cause pneumonia. In adults, bacteria are the most common cause of pneumonia.

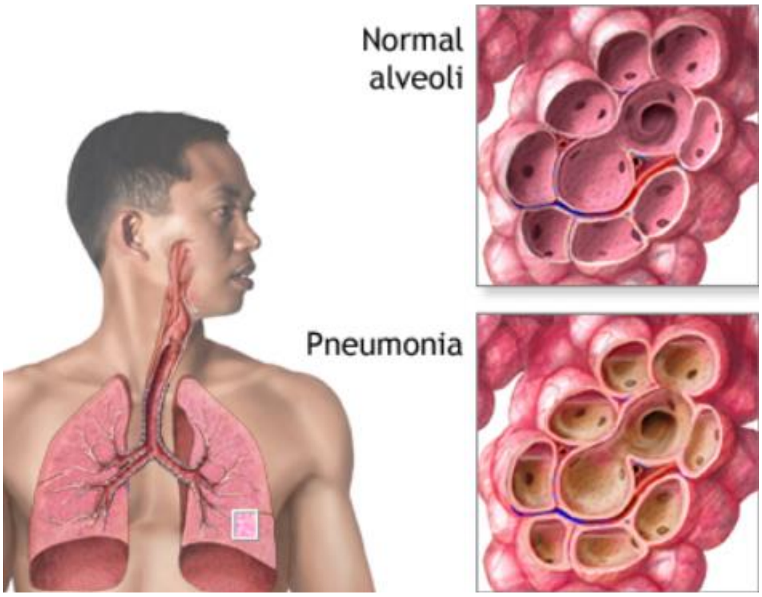


Figure 1.7: Normal alveoli and pneumonia [44]

7- Pulmonary edema is an abnormal buildup of fluid in the lungs. This buildup of fluid leads to shortness of breath.

Pulmonary edema is often caused by congestive heart failure. When the heart is not able to pump efficiently, blood can back up into the blood vessels that take blood through the lungs.

As the pressure in these blood vessels increases, fluid is pushed into the air spaces (alveoli) in the lungs. This fluid reduces normal oxygen movement through the lungs. These two factors combine to cause shortness of breath.[49]

8- Obstructive sleep apnea (OSA): occurs when a loss of muscle tone during sleep in the tongue, soft palate or other soft tissues of the throat allows the airway to collapse and obstructs the flow of air when trying to breathe in. This typically causes a drop in blood oxygen level and a rise in blood carbon dioxide level.

The brain responds with a brief arousal to "jump-start" breathing. This disruption of sleep repeats throughout the night, but most people are not aware of it, because it does not cause them to fully wake up. Even though it may not wake up, the sleep disruption can make people sleepy during the day, no matter how long they sleep at night.[17]

1.8 strategies for healthier respiratory system:

To keep the respiratory system healthy:

- **Avoid** smoke **or** vaping. Smoking causes many lung and airway diseases or makes them worse. Vaping liquids often have many of the same ingredients as cigarettes.
- **Avoid pollutants that can damage airways.** This includes secondhand smoke, chemicals and radon (a radioactive gas that can cause cancer). Wearing a mask when exposing to fumes, dust or other types of pollutants during the job or hobbies.
- **Stay hydrated.** Drinking plenty of water keeps the mucus in lungs thin and easier to clear out.
- **Exercise regularly.** Exercise keeps the muscles in lungs strong and makes breathing easier.
- **Prevent infections.** Washing hands often and getting vaccinated against respiratory illnesses can help prevent from getting sick.[18]

1.9 The time to call a healthcare provider:

Contacting healthcare provider important when having long-lasting or worsening cough, shortness of breath or other symptoms of a respiratory condition. Furthermore, seeing provider for regular checkups. Early diagnosis of respiratory issues can help prevent becoming severe.

1.10 Summary:

In this chapter, we talked about the respiratory system, its contents, its importance, how does it work, the respiratory diseases and their reasons, and the importance of respiratory sounds to determine the respiratory illnesses. In the next chapter we will talk about artificial intelligence and the algorithms we used to build the model that can detect illnesses depending on respiratory sound.

Chapter 2: Artificial Intelligence

2.1 introduction:

Artificial Intelligence is the field of developing computers and robots that are capable of behaving in ways that mimic and go beyond human capabilities, such as decision making, object detection, solving complex problems and so on. AI-enabled programs can analyze and contextualize data to provide information or automatically trigger actions without human interference.[38]

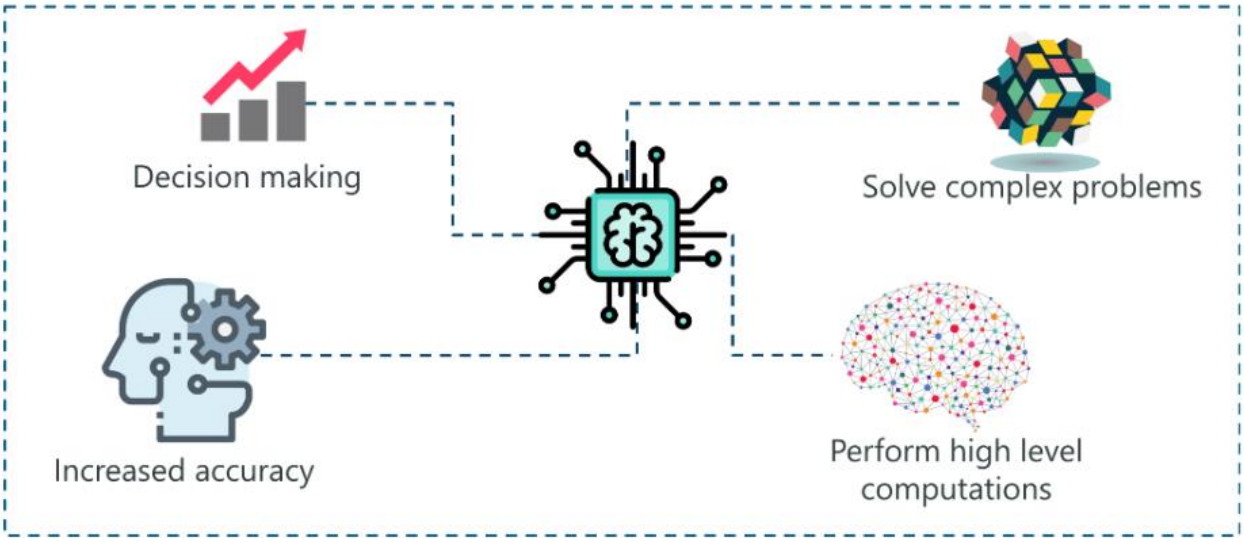


Figure 2.1: Four shapes for AI capabilities [21].

2.2 Stages of Artificial Intelligence:

Artificial intelligence has three main stages: Artificial General Intelligence, Artificial Narrow Intelligence, and Artificial Super Intelligence

2.2.1 Artificial Narrow Intelligence (ANI):

which called weak AI and it is the stage of Artificial Intelligence involving machines that can perform only a narrowly defined set of specific tasks. At this stage, the machine does not possess any thinking ability, it just performs a set of pre-defined functions.

For Example, Siri, Alexa, Self-driving cars, Alpha-Go, Sophia the humanoid. Almost all the AI-based systems built till this date fall under the category of Weak AI.

Weak AI has the possibility to cause harm if a system fails.

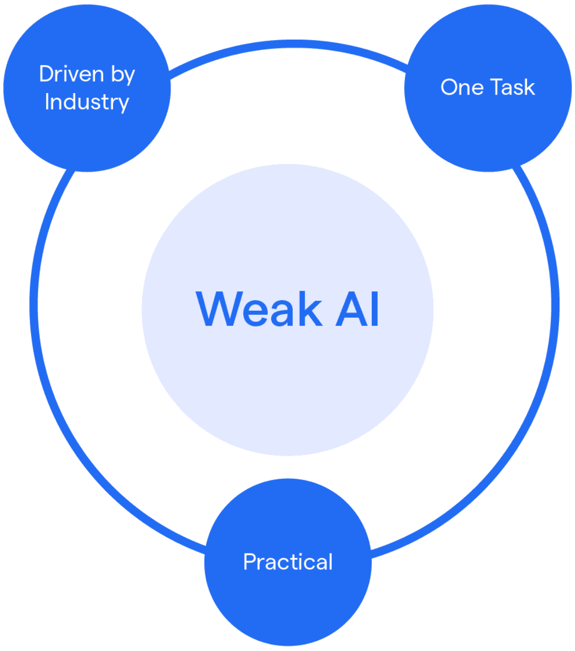


Figure 2.2: Description of weak AI [20].

2.2.2 Artificial General Intelligence (AGI):

It’s called strong AI, and it is the stage in the evolution of Artificial Intelligence wherein machines will possess the ability to think and make decisions just like humans.

Unlike narrower weak AI systems, strong AI would have open-ended abilities to learn, reason, and adapt to unfamiliar environments.

While human input accelerates the growth phase of Strong AI, it develops a human-like consciousness instead of simulating it,

Researchers are still working on creating machines that acts just like human so there are no clear examples for it.[21]

2.2.3 Artificial Super Intelligence (ASI):

is a hypothetical software-based artificial intelligence (AI) system with an intellectual scope beyond human intelligence. At the most fundamental level, this super intelligent AI has cutting-edge cognitive functions and highly developed thinking skills more advanced than any human.

Because ASI can operate continuously, it would be ideal for tasks like safety navigating networks of self-driving cars and assisting in space exploration.

Furthermore, ASI's superior creativity and ability to analyze vast amounts of data might lead to solutions humans can't even imagine, resulting in, hopefully, better quality of life and perhaps even a prolonged life.[20]

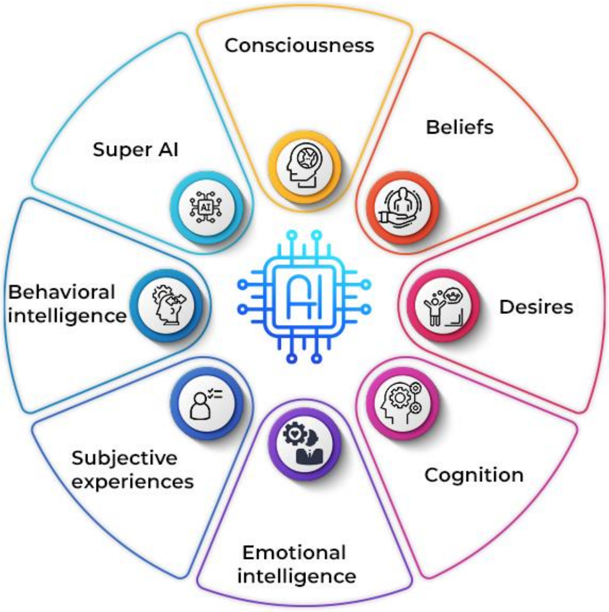


Figure 2.3: Human-like capability of super AI [20].

2.3 Working Mechanism of AI:

Artificial intelligence systems depend on 5 phases appeared in the Figure 2.4:

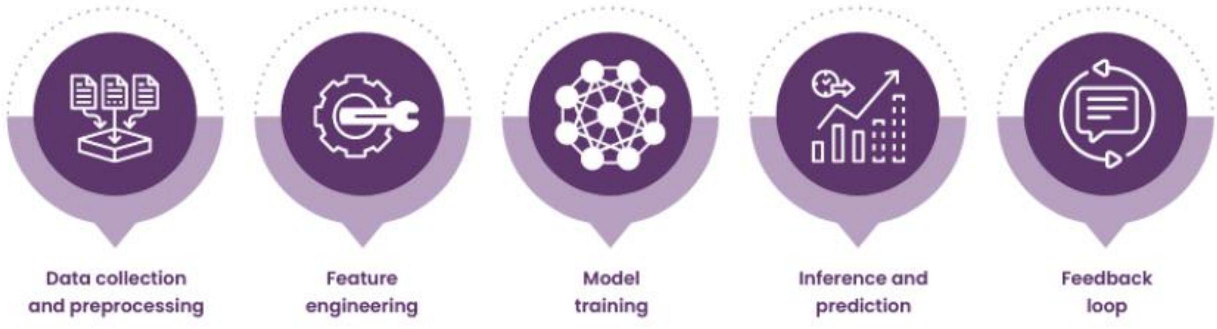


Figure 2.4: Artificial Intelligence mechanism.[20]

2.3.1: Data collection and preprocessing:

Data is the foundation of AI and machines can learn and make decisions based on the data they receive. In the data collection phase, relevant data is gathered from various sources like sensors, databases, or the Internet.

The data may include text, images, videos, or numerical values.

Data often need preprocessing to ensure that it is in a format suitable for analysis, this includes cleaning the data to remove noise or errors, handling missing values, and scaling or normalizing numerical features to bring them into a consistent range.

Data preprocessing is crucial as the data quality directly impacts the performance of AI models.

2.3.2: Feature extraction:

In this phase, we need to select and transform relevant feature carefully from the data to build a predictive model.

Sometimes, we need to create new features through mathematical transformations or by combining existing ones. The goal is to provide the AI model with the most informative and discriminative input variables to make accurate predictions.

2.3.3: Model Training:

In this phase, the AI system can made predictions and decisions from the preprocessed data, and this is where machine learning come into play.

During training, the model iteratively refines its understanding of the data by comparing its predictions to the actual outcomes (labels or target values) in the training dataset. This process is often guided by a loss or cost function that quantifies the model's performance. The goal is to minimize this function to achieve the most accurate predictions and find accuracy.

2.3.4: Inference and prediction:

After training the AI model, it's ready for Inference and Prediction.

In this phase, the model uses the knowledge gained during training to make predictions or decisions on new, unseen data. It applies the learned patterns and relationships to the input data, and based on these patterns, it provides predictions or classifications. It can be in the real time or on a punch of data.

2.3.5: Feedback loop:

The feedback loop represents an essential aspect of AI systems.

Therefore, we can continuously improve the model's performance based on new data and user feedback.

As the AI system operates in the real world, it collects new data, which can be used to retrain the model periodically. The feedback loop ensures that the AI system remains up-to-date and adapts to changing conditions.

2.4 Domains of Artificial Intelligence:

Artificial Intelligence is a diverse field that contains various algorithms and techniques. These algorithms and techniques are the building blocks of AI systems, each serving a unique purpose. It includes sex fundamental domains:

2.4.1: Machine Learning:

Machine learning is a field of artificial intelligence that allows computers to learn from large datasets by identifying patterns and relationships within data.

It uses historical data as input to predict new output values. Classical (non-deep) machine learning models require more human intervention to segment data into categories.

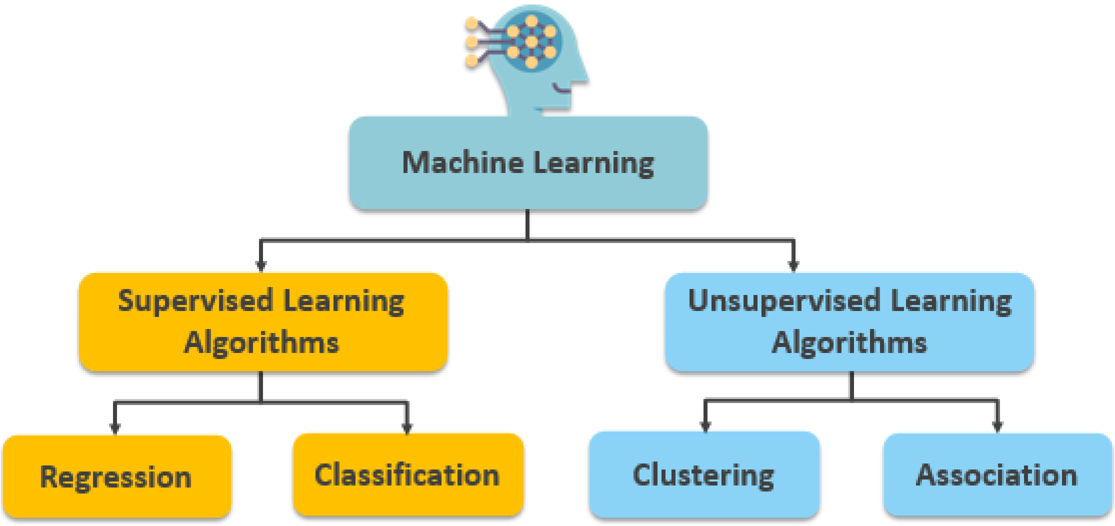

Machine Learning
Supervised Learning Algorithms
Unsupervised Learning Algorithms
Regression
Classification
Clustering
Association

Figure 2.5: Supervised and unsupervised learning branches. [38]

Machine learning consists of three main categories:

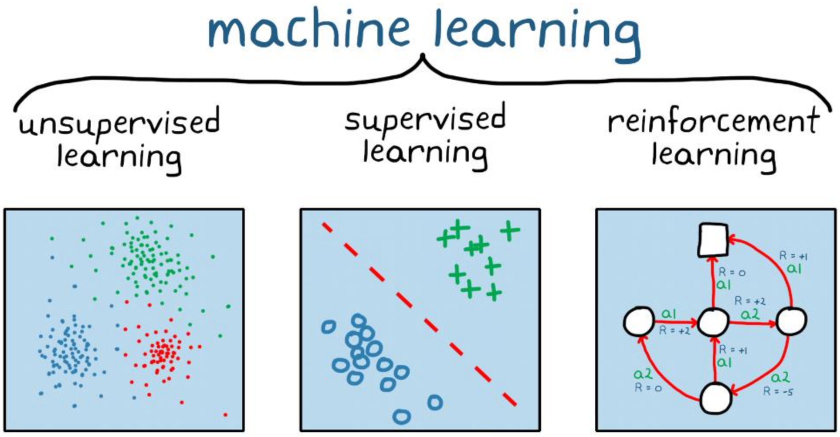


Figure 2.6: Three types of Machine learning. [22]

- **Supervised learning:** it’s the most common and widely used machine learning techniques. Through this category, the model is trained using a labelled dataset, where each input is paired with the correct output.
  The algorithm learns to configure inputs to outputs by identifying patterns and relationships in the data.
  Classification: is a vital aspect of ML, that involves assigning unknown data points to predefined classes.
  Regression: is also a vital aspect of ML, where we aim to predict real values based on the features present in the training data.
  Unlike classification, where we assign classes, regression focuses on estimating continuous and numerical values.

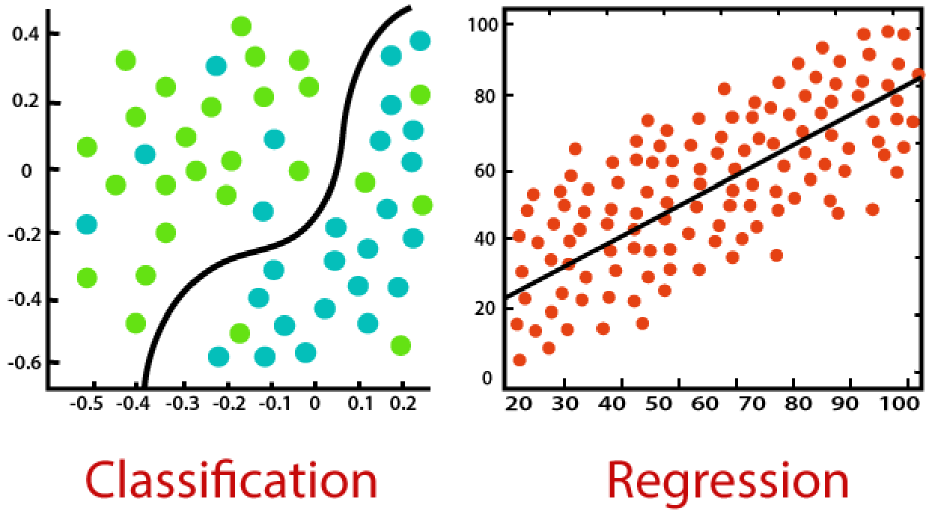


Figure 2.7: The difference between classification and regression. [23]

- **Unsupervised learning:** Unsupervised Learning is when the algorithm identifies patterns and structures by clustering similar data points together or reducing the dimensionality of data.
  clustering does not involve assigning predefined labels or predicting values. Instead, clustering groups data points based on the similarity of their features. The goal is to identify inherent patterns and relationships within the data. Clustering is particularly useful when we have unlabeled data and want to discover hidden structures.

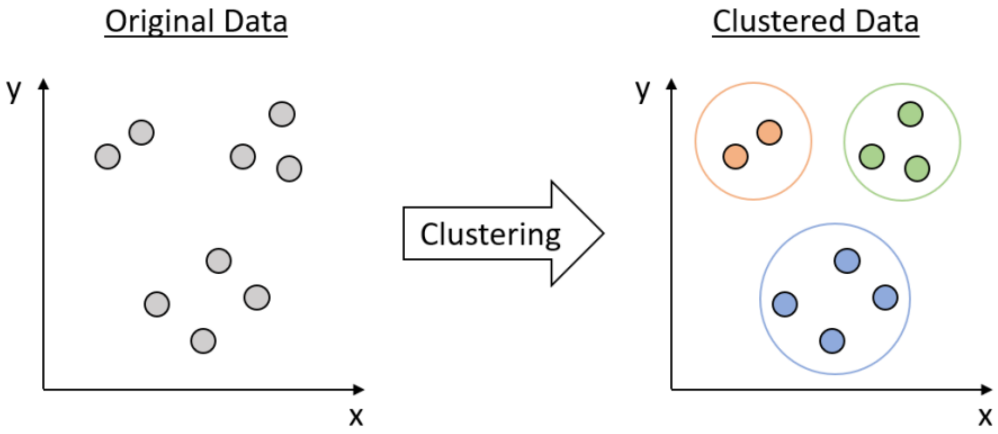


Figure 2.8: definition of data clustering. [38]

- **Reinforcement Learning:** is a machine learning (ML) technique that trains software to make decisions to achieve the most optimal results. It mimics the trial-and-error learning process that humans use to achieve their goals.

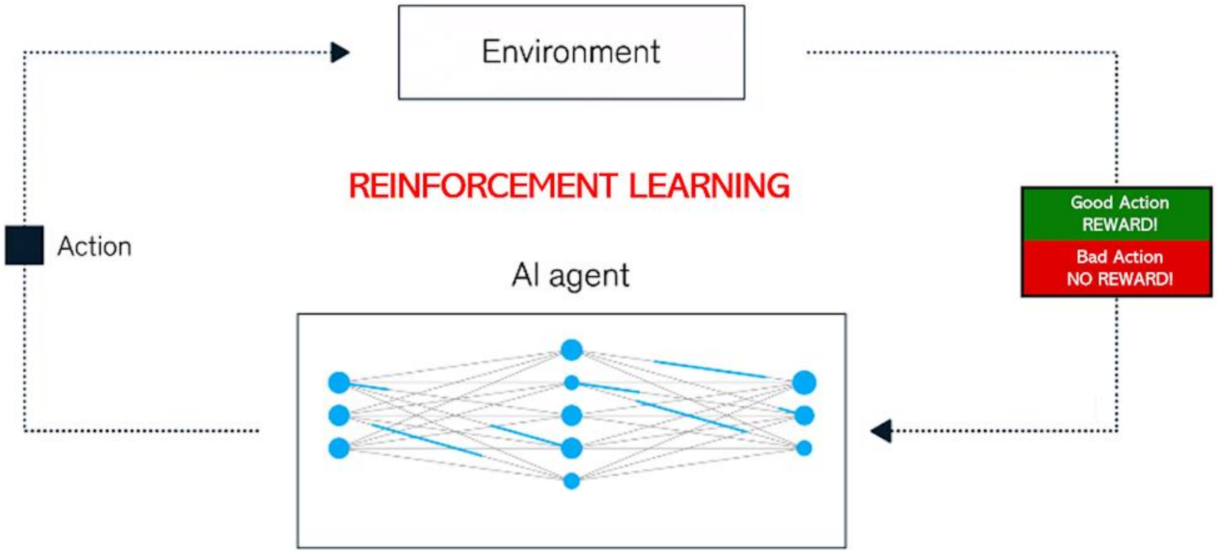


Figure 2.9: Reinforcement learning. [34]

2.4.2: Natural Language Processing:

Natural language processing (NLP) is the science of training computers to understand and produce written and spoken language in a similar manner as humans, as its primary input is usually text or voice data.

Unlike traditional machine learning models, which might incorporate features such as numerical and categorical data alongside text, NLP-based models typically concentrate directly on analyzing and understanding the patterns within textual or spoken language for tasks such as analysis and prediction.

Twitter uses NLP to filter out terroristic language in their tweets, and Amazon uses NLP to understand customer reviews and improve user experience.

NLP mainly tackles speech recognition and natural language generation, and it’s leveraged for use cases like spam detection and virtual assistants.

2.4.3: Robotics:

In Robotics, we integrate AI and machine systems to create machines that can do human physical tasks and interactions with the environment.

It enables autonomous and semi-autonomous robots to perform various functions in industries like manufacturing, healthcare, and exploration new versions of robotics now, to allow them to take decisions and interact with the environment around them.

One of the best examples of robotics is Sofia.

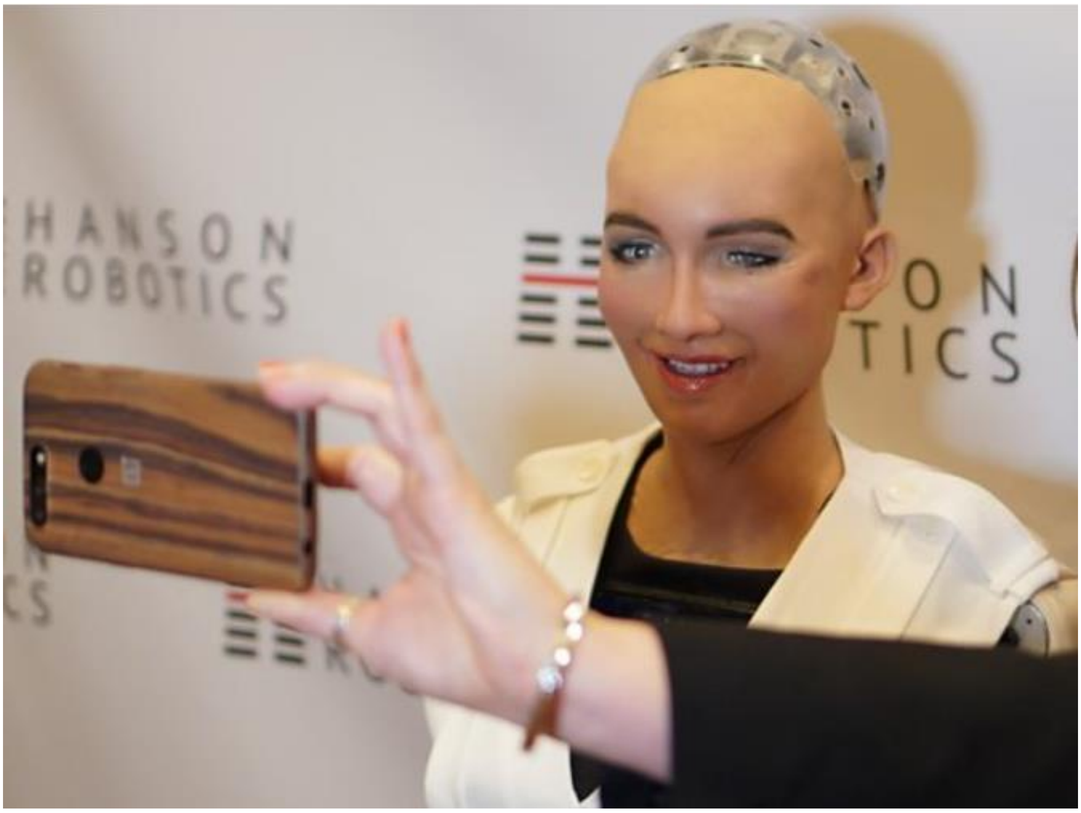


Figure 2.10: Robot “Sofia” [21]

2.4.4: Computer vision:

Computer vision, is a branch of artificial intelligence that enables machines to understand visual information in the world, much like how humans perceive and analyze images and videos. Therefore, the input of computer vision is images or videos in real time or previously recorded.

Unlike traditional machine learning models, which might incorporate features such as numerical and categorical data alongside images and videos, Computer Vision models typically concentrate directly on analyzing and understanding the patterns within image or video data for tasks like object recognition, segmentation, and scene understanding.

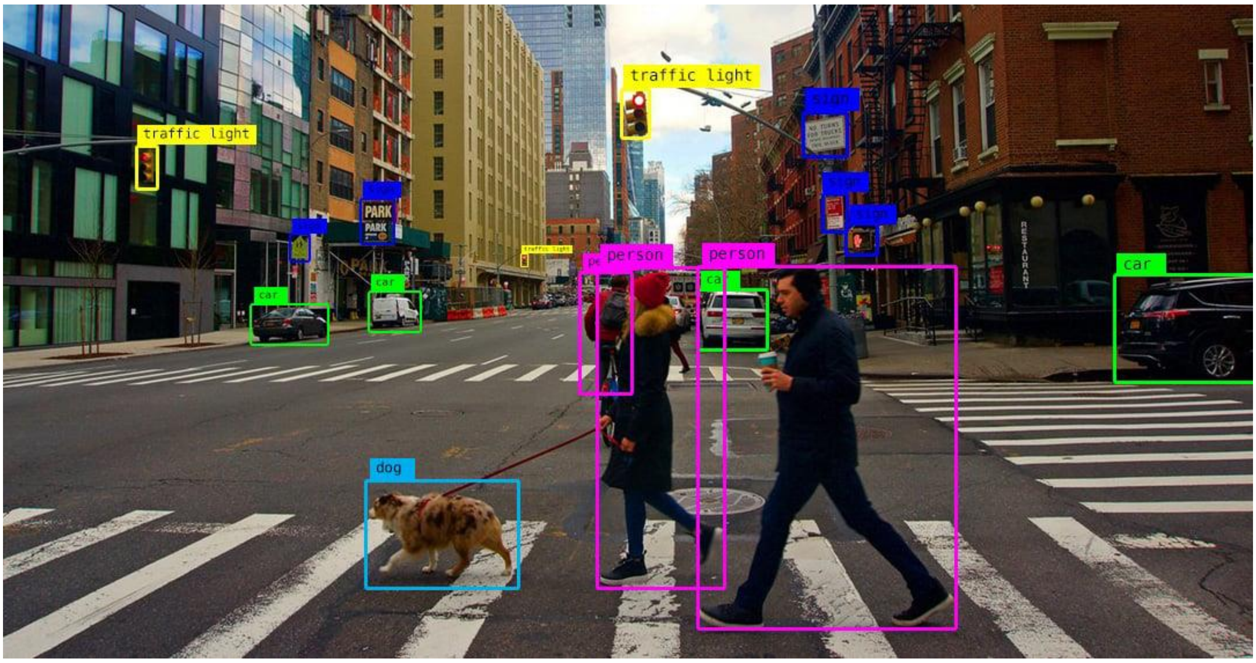


Figure 2.11: Recognizing objects using computer vision.[28]

2.4.5: Espert system:

Expert systems are AI-based computer systems that aim to mimic the knowledge and expertise of human experts in a specific domain. They use rules and logic-based reasoning to solve complex problems and provide expert-level advice.

Expert systems are mainly used in information management, medical facilities, loan analysis, virus detection and so on.

2.4.6: Deep Learning:

Deep Learning is the process of implementing Neural Networks on high dimensional data to gain insights and form solutions. These networks can automatically learn intricate hierarchical representations from data, making them suitable for complex tasks.

Neural networks are the backbone of deep learning algorithms, they are called “neural” because they mimic how neurons in the brain signal one another.

A neural network consists of neurons (perceptron) interconnected like a web and these neurons are mathematical functions or models that do the computations required for classification according to a given set of rules.

Deep neural network consists of the following layers:

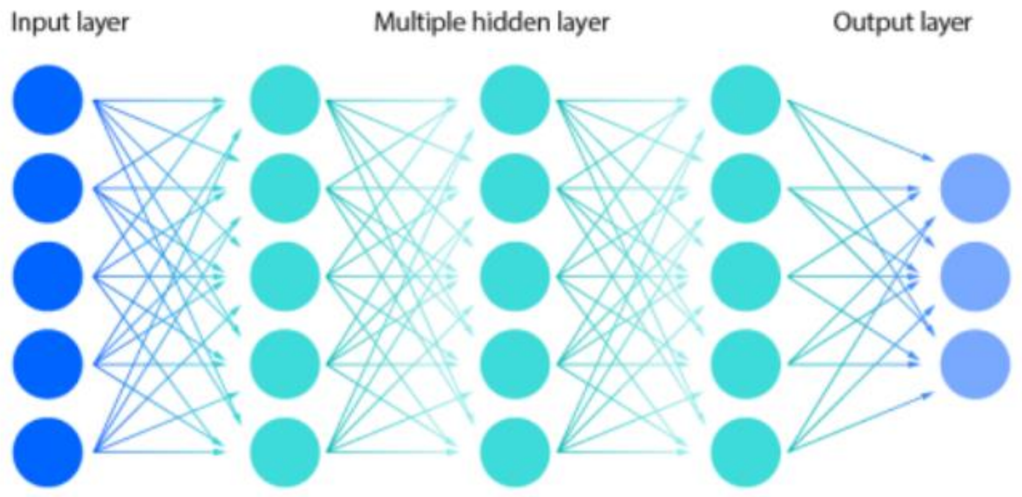


Figure 2.12: Neural networks layers [22]

- Input layer

An artificial neural network has several nodes that input data into it. These nodes make up the input layer of the system.

- Hidden layer

the data is passed from the input layer to the hidden layer after it is processed well, then the hidden layers process information at different levels, adapting their behavior as they receive new information. Deep learning networks have hundreds of hidden layers that they can use to analyze a problem from several different angles.

For example, if we were given an image of an unknown animal that we had to classify, we would compare it with animals we already know. For example, we would look at the shape of its eyes and ears, its size, the number of legs, and its fur pattern.

If a deep learning algorithm is trying to classify an animal image, each of its hidden layers processes a different feature of the animal and tries to accurately categorize it.

- Output layer

The output layer consists of the nodes that output the data. Deep learning models that output "yes" or "no" answers have only two nodes in the output layer. On the other hand, those that output a wider range of answers have more nodes.

2.4.6.1 The steps of working neural networks:

Neural networks are computational models inspired by the human brain. They consist of interconnected layers of neurons (nodes) that process data and learn patterns. Here's how they work with formulas:

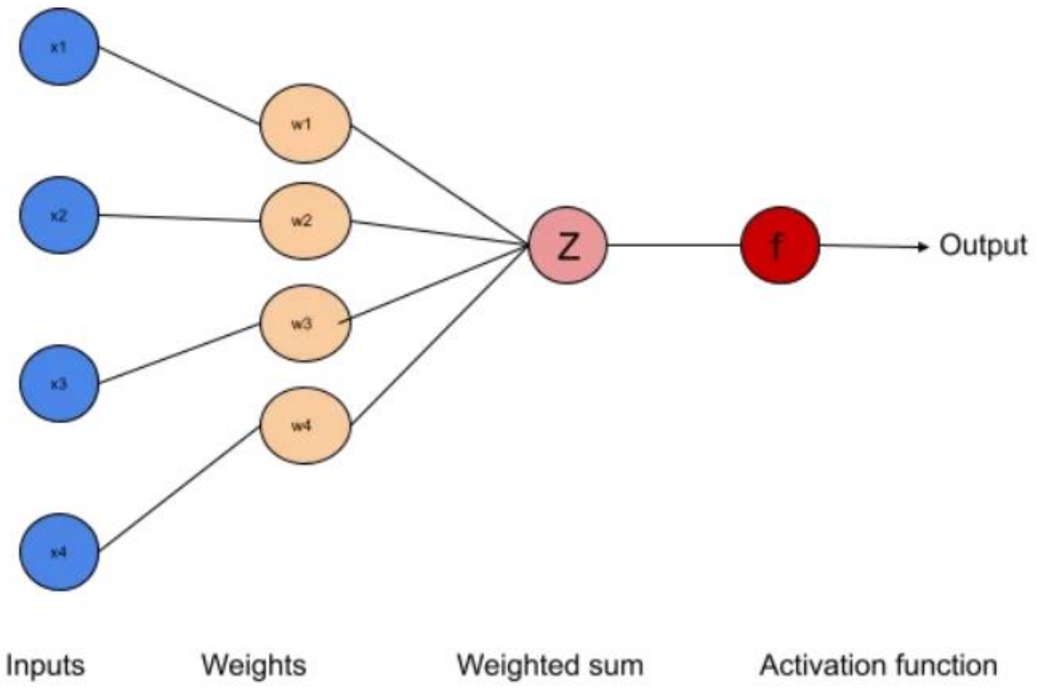


Figure 2.13: One perceptron components [40]

1. Input Layer: The input layer receives the input data. Each input is represented as x_i where i ranges over the number of inputs.

2. Weights and Biases: Each connection between neurons has a weight w_{ij} and each neuron has a bias b_j.

3. Neuron Activation:

$$z_j = \sum_i w_{ij} x_i + b_j \qquad (2.1)$$

Here, z_j is the weighted sum of inputs for neuron j.

4. Activation Function: This function introduces non-linearity to the network. Common activation functions include:

The weighted sum and bias are passed through an activation function, determining neuron activation.

- Sigmoid:

$$\sigma(z) = \frac{1}{1+e^{-z}} \qquad (2.2)$$

- ReLU (Rectified Linear Unit):

$$\mathrm{ReLU}(z) = \max(0, z) \qquad (2.3)$$

5. Feature Extraction: Neurons fire based on the activation function output, enabling feature extraction.

Output: $a_j = activation\ z_j$ (2.4)

6. Loss Function: The loss function quantifies the difference between the predicted output and the actual output. For example, in classification, cross-entropy loss is used:

$$(2.5) \qquad L = -\sum_i y_i \log(\hat{y}_i)$$

7. Backpropagation: The network learns by updating weights and biases to minimize the loss. This is done through backpropagation:

$$(2.6) \qquad \frac{\partial L}{\partial w_{ij}} = \frac{\partial L}{\partial z_j} \cdot \frac{\partial z_j}{\partial w_{ij}}$$

Using gradient descent, weights are updated:

$$(2.7) \qquad w_{ij} \leftarrow w_{ij} - \eta \frac{\partial L}{\partial w_{ij}}$$

Here, η is the learning rate.

Weights are adjusted to minimize error through backpropagation, the central learning mechanism.

8. Iterations (Epochs): This process of forward pass, loss calculation, backpropagation, and weight update are repeated over multiple iterations (epochs) until the network converges to a solution.

Therefore, training a neural network involves iteratively calculating the error, adjusting weights and biases through backpropagation, and fine-tuning the model parameters until the error is within an acceptable range.

2.4.6.2: Deep neural networks types:

- Artificial neural networks (ANN): is a group of multiple neurons at each layer. In ANN, input is only processed in the forward direction through various input nodes until it makes it to the output node, and it's one of the simplest variants of neural networks.
  Advantages of ANN: ability to work with incomplete knowledge, have fault tolerance, have a distributed memory.
  Disadvantages of ANN: hardware dependance, unexplained behavior of the network, and determination of paper network structure.
  ANN is used in image recognition, predictive modeling, natural language processing (NLP), autonomously flying aircraft, detecting credit card fraud, mastering the game of Go, and many more.

- Recurrent neural networks (RNN): save the output of the processing nodes and feed the results back into the model and they pass the information in both directions. Each node in RNN model acts as a memory cell, continuing the computation and implementation of operations.
  RNN has the ability to self-learn and continue working toward the correct direction during backpropagation if the system prediction is incorrect.

  Advantages of RNN: ability to remember every information through time so that it's useful for time series prediction and this is what we call long short-term memory (LSTM).
  Disadvantages of RNN: training RNN is very difficult task, and it cannot process very long sequences if using tanh as an activation function

  RNN is widely used for speech recognition, natural language processing (NLP), machine translation and time series forecasting.

- Conventional neural networks (CNN): it’s the most popular model used nowadays, it uses a variation of multi-layer neurons and contains one or more conventional layers that can be either entirely connected or pooled.

  Advantages of CNN:
  CNN has very high accuracy in image and audio recognition problems, the ability to weight sharing and can detect the important features without human interference.
  Disadvantages of CNN:
  It doesn’t encode the position and orientation of objects, weak in the spatial data, and lots of training dataset is required.

  CNNs are similar to feedforward networks, but they’re usually utilized for image recognition, audio detection, pattern recognition, and/or computer vision. These networks harness principles from linear algebra, particularly matrix multiplication, to identify patterns within an image.

2.5 Transformers:

A transformer model is a neural network that learn context and thus meaning by tracking relationships in sequential data like the words in this sentence.
Transformers are translating text and speech in near real-time, opening meetings and classrooms to diverse and hearing-impaired attendees.[24]

2.5.1 Architecture of Transformers:

The transformer consists of an encoder and a decoder, each composed of multiple layers. However, many transformer applications (like BERT or GPT) use only the encoder or the decoder.

- **Encoder**: The encoder processes the input sequence and creates a set of continuous representations.
- **Decoder**: The decoder uses the encoder's output and the previously generated tokens to produce the output sequence.[24]

2.5.2 Key component of transformer:

The main key components of the transformers summarized in the following:

a. Multi-Head Self-Attention Mechanism

- **Self-Attention**: Self-attention allows each position in the input sequence to attend to all other positions, helping the model weigh the importance of each token relative to others. In the sentence example: each word in a sequence to interact with all other words, capturing dependencies and relationships regardless of their distance in the sequence in order to enhance the model's ability to understand complex sentences.

  This is done using three matrices: Query (Q), Key (K), and Value (V):

$$Attention\ (Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_K}}\right)V$$

Where:

Queries (Q): Derived from the input sequence.

Keys (K): Also derived from the input sequence.

Values (V): Same as the keys, but represent the actual values to be attended to.

Scaled Dot-Product Attention: The dot products of the query with all keys are computed, scaled by the dimension of the keys, and passed through a SoftMax function to obtain the weights on the values.

- **Multi-Head Attention**: Instead of performing a single attention function, the transformer uses multiple attention heads to capture different aspects of the relationships between words.

$$MultiHead(Q,K,V) = concat\ (head_1, head_2, \dots \dots .\ head_H)W^O$$

b. Positional Encoding:

Since transformers do not have a built-in sense of order (like RNNs or CNNs), positional encodings provide information about the position of each word in a sequence, which is essential for transformer models that process input in parallel. They use sine and cosine functions to add this positional information to the input embeddings.

c. Feed-Forward Neural Network

Each encoder and decoder layer contains a position-wise fully connected feed-forward network applied to each position separately and identically.

$$FFN(x) = \max(0, xW_1 + b_1)\,W_2 + b_2; \quad (2.9)$$

W is the weight, b is the bias

The encoder consists of Multi-Head Self-Attention Mechanism, Feed-Forward Neural Network, and Residual Connections and Layer Normalization applied to both the attention and feed-forward sub-layers.

While decoder consists of Masked Multi-Head Self-Attention Which Prevents attending to future tokens in the sequence, Multi-Head Attention which attends to the encoder's output, Feed-Forward Neural Network, and Residual Connections and Layer Normalization similar to the encoder.[26]

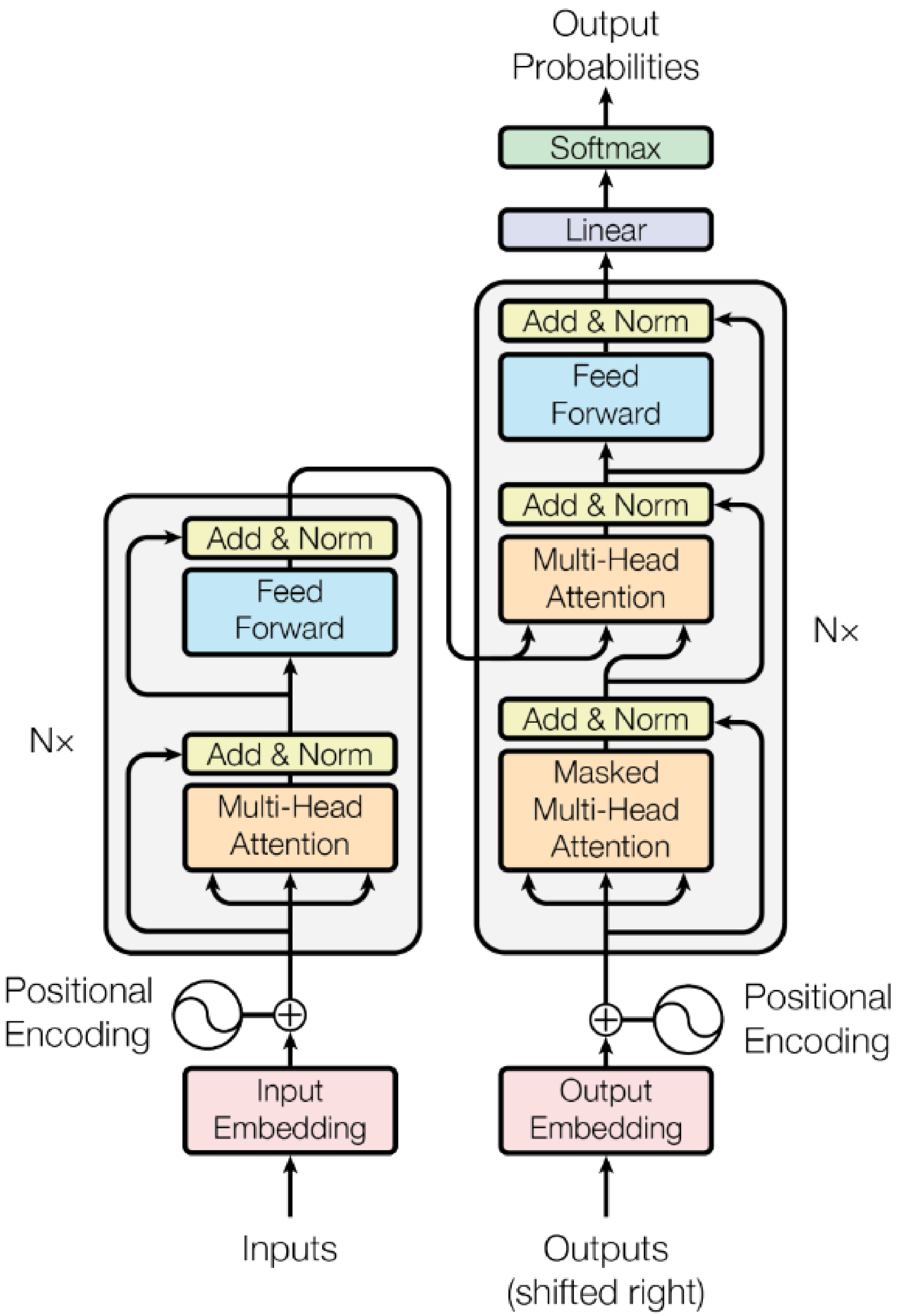


Figure 2.14: The transformer model architecture [26]

2.5.3 Mechanism of Transformer:

Input Embedding: The input tokens are converted into embeddings and positional encodings are added.

Encoding: The embeddings are passed through the encoder layers. Each layer applies multi-head self-attention and feed-forward networks.

Decoding: The decoder takes the encoder's output and the previously generated tokens to produce the next token. This process continues until the end of the sequence is generated. [54]

2.5.4 Advantages of Transformers:

1- **Parallelization**: Unlike RNNs, transformers do not require sequential processing, allowing for parallelization and faster training.
2- **Long-Range Dependencies**: Self-attention mechanisms allow transformers to capture long-range dependencies more effectively than RNNs.
3- **Flexibility**: Transformers have been adapted to various tasks beyond NLP, including image and speech processing.

2.6 Summary:

In this chapter, we defined AI, its subtitles, machine learning, deep learning, supervised and unsupervised learning. We also referred to the most important algorithms that we used (CNN, RNN, and transformer). In the next chapter, we will talk about the previous studies and how we will add our value.

Chapter 3 - Reference Studies

3.1 First Study:

A Respiratory Sound Database for the Development of Automated Classification

A study conducted by Rocha *et al*. at the University of Coimbra, Portugal (2017) [53], aimed to develop algorithms capable of characterizing respiratory sound recordings from clinical and non-clinical environments, thereby advancing diagnostic capabilities in respiratory health assessment. To achieve this, the researchers designed a deep learning algorithm to detect respiratory sounds, including crackles and wheezes. Crackles are discontinuous, explosive, and non-musical adventitious sounds commonly associated with cardiorespiratory diseases, while wheezes are musical sounds lasting more than 250ms, often observed in patients with obstructive airway diseases such as asthma and COPD.

The dataset, collaboratively created by research teams in Portugal and Greece, comprised 920 recordings from 126 subjects, totaling 6898 respiration cycles over 5.5 hours. Respiratory experts annotated the cycles for crackles, wheezes, a combination of both, or no adventitious sounds. Recordings, collected using diverse equipment, varied in duration from 10s to 90s, and included chest location information. Despite the presence of high levels of noise mirroring real-life scenarios, data augmentation techniques were employed, with one-hot encoding preferred for labeling.

The researchers implemented a sophisticated convolutional neural network (CNN) tailored to analyze Mel-Spectrograms extracted from the respiratory sound recordings. This CNN architecture facilitated automatic identification and classification of crackles and wheezes by leveraging distinctive patterns within the spectrograms. Despite these efforts, classifying 'wheeze' and 'wheeze and crackles' instances remained challenging, often resulting in false negatives and suboptimal recall scores. The overall validation accuracy currently stands at approximately 70%.

The study underscores the significant potential of deep learning techniques, such as CNNs, in respiratory sound analysis, with implications for advancing diagnostic capabilities and patient care.[3]

3.2 Second Study:

Deep Neural Network for Respiratory Sound Classification in Wearable Devices Enabled by Patient Specific Model Tuning.

The study was conducted by Acharya and Basu at the University of Illinois at Urbana-Champaign, United States (2020) [9]. The researchers aimed to develop advanced classification models for identifying breathing sound anomalies, such as wheezes and crackles, to automate the diagnosis of respiratory and pulmonary diseases.

To achieve this, the researchers proposed a hybrid CNN-RNN model that classifies respiratory sounds using Mel-spectrograms. They introduced a patient-specific model tuning strategy that initially screens respiratory patients and subsequently builds tailored classification models for reliable anomaly detection using limited patient data. Additionally, they implemented a local log quantization technique to reduce the memory footprint of the model, making it suitable for deployment in memory-constrained systems such as wearable devices.

The hybrid CNN-RNN model achieved a score of 66.31% on the four-class classification of breathing cycles in the ICBHI'17 scientific challenge respiratory sound database. When re-trained with patient-specific data, the model scored 71.81% in leave-one-out validation. The proposed weight quantization technique resulted in approximately a 4× reduction in total memory cost without performance loss.

The main contributions of this study are threefold: achieving state-of-the-art scores on the ICBHI'17 dataset, demonstrating that deep learning models can successfully learn domain-specific knowledge and outperform generalized models, and showing that local log quantization of trained weights can significantly reduce memory requirements. This patient-specific re-training strategy is valuable for developing reliable, long-term automated patient monitoring systems, particularly in wearable healthcare solutions.

3.3 Third Study:

A Neural Network-Based Method for Respiratory Sound Analysis and Lung Disease Detection

A study conducted by Grzywalski et al. at the University of Molis, Italy (2022) [4]. The study aimed to develop neural network models for the classification of respiratory sounds to aid in the detection of lung diseases. The researchers focused on enhancing diagnostic processes by applying deep learning techniques to analyze respiratory sound recordings.

The study utilized a comprehensive dataset consisting of respiratory sound recordings, which included various respiratory conditions such as crackles and wheezes. These sounds are critical indicators of underlying respiratory diseases, with crackles being short, explosive sounds typically associated with conditions like pneumonia and fibrosis, and wheezes being continuous musical sounds linked to obstructive diseases like asthma and COPD.

To achieve accurate classification, the researchers tried four different models to get the best result: KNN, SVM, Random Forest and a convolutional neural network (CNN) architecture. The CNN is well-suited for analyzing audio data represented as spectrograms. modes were trained to distinguish between normal and abnormal respiratory sounds, including specific classifications for crackles and wheezes.

The results demonstrated that the CNN model achieved the best accuracy in classifying respiratory sounds. As it got the best classification result which is 91.2%, showcasing its potential for clinical applications in respiratory disease diagnosis. While the accuracy for SVM is 87.5%, for KNN is 84.3%, and for Random Forest is 86%

The study approved the effectiveness of deep learning models in processing and interpreting respiratory sounds, and it’s a very crucial tool for early detection and diagnosis of lung diseases. This approach can greatly enhance patient health by enabling timely and accurate intervention based on automated sound analysis.

3.4 Summary:

We noticed in this chapter that each study got different accuracy, the first one focused on classifying crackles and wheezes and CNN showed the best accuracy: 70%. The second classified respiratory illnesses and got the best accuracy from CNN: 71.81%. the third one got the best accuracy also from SVM which is 87.5%. in our study, we try CNN, RNN, and compare them with Bert transformer which is a new algorithm that was used to build AI APPs such as ChatGpt and Gemeni. We also want to determine the best features to be used for audio data that can extract the best information.

Aim of Study

The aim of the study is to find algorithms with high accuracy and sensitivity capable of predicting respiratory diseases prognosis depending on respiratory sounds to help early, harmless and non-invasive diagnosis in order to be able to intervene quickly in the patient's treatment protocol to reduce mortality as much as possible.

Practical Review

Chapter 4 – Methods and Materials

4.1. Introduction:

Diagnosing respiratory diseases using respiratory sounds is an effective, accessible, and efficient approach that enhances early detection, continuous monitoring, and overall patient care. It leverages simple, non-invasive techniques to provide rich diagnostic information and supports the integration of technology in healthcare, thereby improving outcomes for patients with respiratory conditions.

Here are some advantages of using respiratory sounds to detect respiratory illnesses:

- **Noninvasive and safe:** as it's non-invasive method, it does not require any incisions, injections, or exposure to radiation. This makes it a safer and more comfortable option for patients. In addition to the ability of safe repeating it multiple times for continuous monitoring.
- **Cost effective:** the tools used for recording and analyzing respiratory sounds like stethoscopes and digital recorders are way cheaper than diagnostic tools like CT or MRI.
- **Early detection:** regular monitoring of the respiratory system helps in early detection of chronic respiratory conditions like asthma, COPD, and bronchitis. Thus, the cost of cure will be less compared to the cost of cure when late detection.
- **Accessibility:** analyzing respiratory sounds can be performed using portable devices, making it accessible in remote or underserved areas. As well as supporting telemedicine, which is particularly useful in pandemics or for patients with mobility issues.
- **Real-time analysis:** Devices equipped with sound analysis algorithms can provide real-time feedback, which is crucial during emergencies or acute exacerbations of respiratory conditions. [56]

4.2 Dataset:

The data is an open-source data set that was compiled by collaborative efforts from two research teams based in Portugal and Greece in 2017. The dataset includes a total of 920 recordings obtained from 126 subjects, encompassing a broad spectrum of respiratory conditions. These recordings collectively contain 6898 respiratory cycles, each annotated by respiratory experts. The duration of each record varies between 10 to 90 seconds, thereby offering both short and extended samples for analysis. On the other hand, the total duration of the recordings amounts to approximately 5.5 hours. Out of the 6898 respiratory cycles: 1864 cycles contain crackles, 886 cycles contain wheezes, and 506 cycles contain both crackles and wheezes

The annotations classify the cycles as containing crackles, wheezes, a combination of both, or no adventitious respiratory sounds at all.[56]

The team collected sounds from seven chest locations: trachea; left and right anterior, posterior, and lateral. Sounds were collected in clinical and non-clinical (home) settings. The acquisition of RS was performed on subjects of all ages, from infants to adults and elderly people. Subjects included patients with lower respiratory tract infections, upper respiratory tract infections, COPD, asthma, and bronchiectasis.

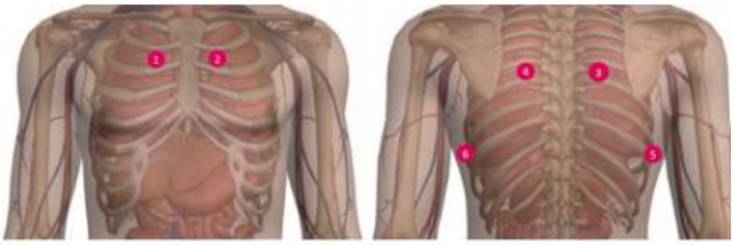


Figure 4.1: Chest locations for recording respiratory sounds [56]

In some studies, the sounds were collected sequentially with a digital stethoscope (Welch Allyn Master Elite Plus Stethoscope Model 5079-400). In other studies, the sounds were collected using either seven stethoscopes (3M Littmann Classic II SE) with a microphone in the main tube or seven air-coupled electret microphones (C 417 PP, AKG Acoustics) located into capsules made of Teflon.

Respiratory sounds were annotated using the Computerized Lung Auscultation – Sound System (CLASS).

The diversity in recording equipment and conditions, including high noise levels in some cycles, simulates real-life scenarios, thus presenting a realistic and challenging environment for sound classification algorithms.

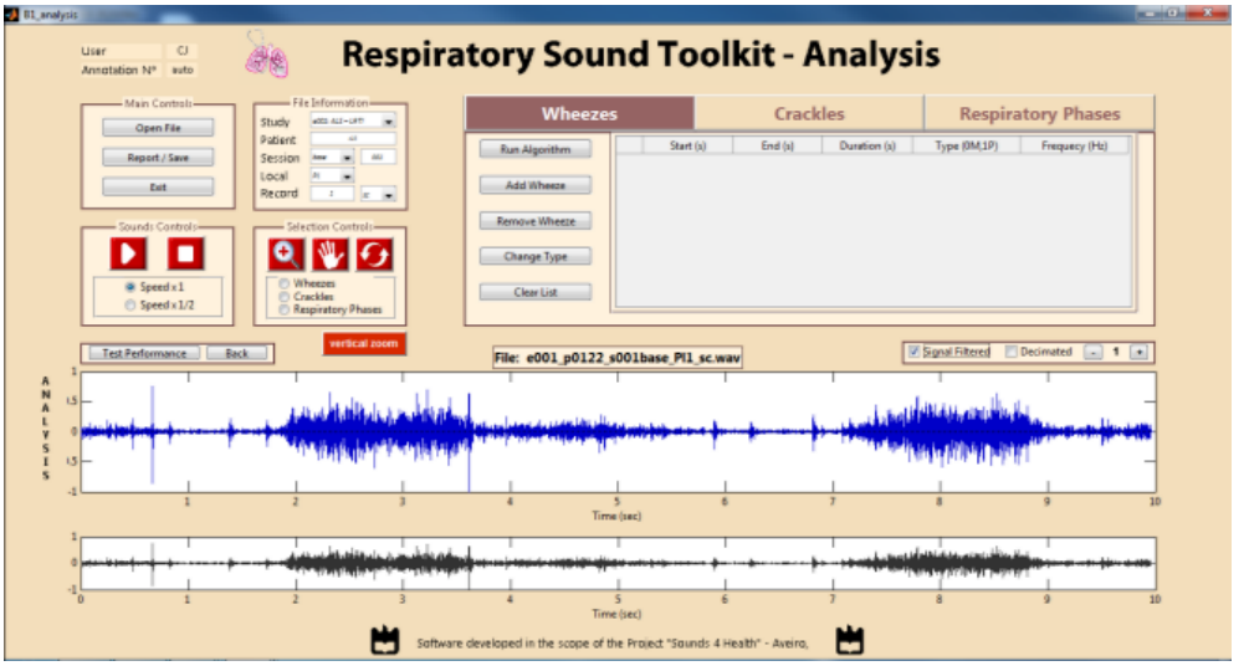


Figure 4.2: The respiratory sound annotation software [56]

4.3 Materials:

There are multiple materials used to achieve the project, starting with the tools used to acquire respiratory sound data, to the tools that used for processing it and extracting results.

4.3.1 Digital Stethoscope:

Digital stethoscopes are tools used by healthcare providers to hear the sounds made by the lungs, heart, and other internal areas, such as the intestinal tract. With digital stethoscopes. Acoustic sounds are converted into electronic signals, helping to amplify the sounds so the healthcare provider can listen with accurately. Additionally, the signals can be transmitted to a computer for additional processing.[9]

There are multi types of digital stethoscope that can be used for acquiring respiratory sounds dataset but we will explain about the two used in our data:

- ✓ Welch Allyn Master Elite Plus Stethoscope Model:

This electronic stethoscope can be connected to the audio-in jack of the multimedia board as the audio input source. It can also be connected to PC, has 200 hours battery life, adjustable ear-piece and a safety cut-off of

95 dB. It also has volume control and two mode settings for heart and lung sounds. Unlike other electronic stethoscopes, this model uses piezoelectric sensors instead of microphones in the chest piece of the stethoscope to eliminate the interference of ambient noise. [8]

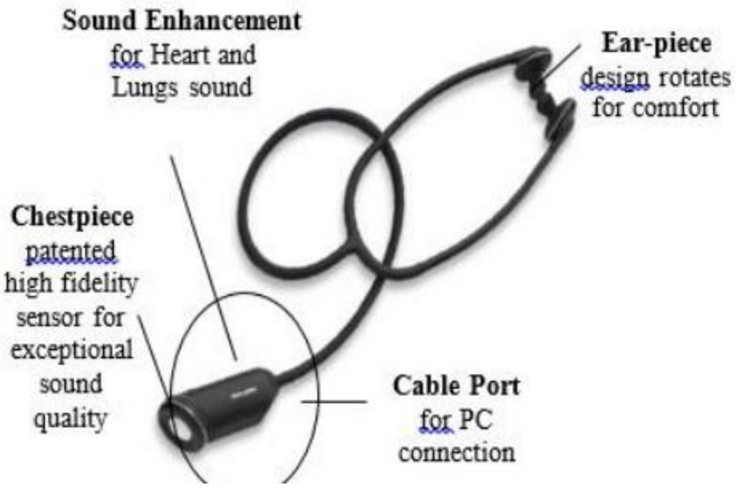


Figure 4.3: Welch Allyn Master Elite Plus Stethoscope [8]

- ✓ seven stethoscopes (3M Littmann Classic II SE): This stethoscope is specially designed for infants, where accuracy and gentleness are paramount. It delivers high acoustic sensitivity, through its dual-sided chest piece, for both high and low frequency sounds. It is easily adjusted for head size and comfort by squeezing together or pulling apart the ear tubes. Snap-tight, soft-sealing ear tips conform to individual ears for an excellent acoustic seal and comfortable fit.[51]

Figure 4.4: 3M Littmann Classic stethoscope [51]

- ✓ seven air-coupled electret microphones (C 417 PP, AKG Acoustics):

The C417 is a professional lavalier microphone with omnidirectional polar pattern, its sound is extremely open and natural, making it ideal for wireless or hardwire multi-mic situations.[25]

It has the following features:

- Omnidirectional polar pattern
- Extremely lightweight and inconspicuous
- Professional three-pin, mini XLR connector
- Balanced sound characteristics

Figure 4.5: The C417 PP Microphone [25]

4.3.2: Data processing tools:

✓ Google Collab: Collab is a hosted Jupyter Notebook service, and it was chosen as it requires no setup to use and provides free of charge access to computing resources, including GPUs and TPUs. It is also especially well-suited to machine learning, data science, and education. Collab allows to use and share Jupyter notebooks with others without having to download, install, or run anything.

Figure 4.6: Google collab [58].

✓ Python: Python is an interpreted, object-oriented, high-level programming language with dynamic semantics that was created by Guido van Rossum, and released in 1991. It works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc). It was used to process data and build AI models because it is the best language for these uses (data analysis, data science) as it’s simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance.

Figure 4.7: Python Language [59].

Python supports modules and packages, which encourages program modularity and code reuse. Python is the language that was used to build new AI-App like ChatGpt, Gemeni …etc, that was I used it.

4.4 Practical Steps:

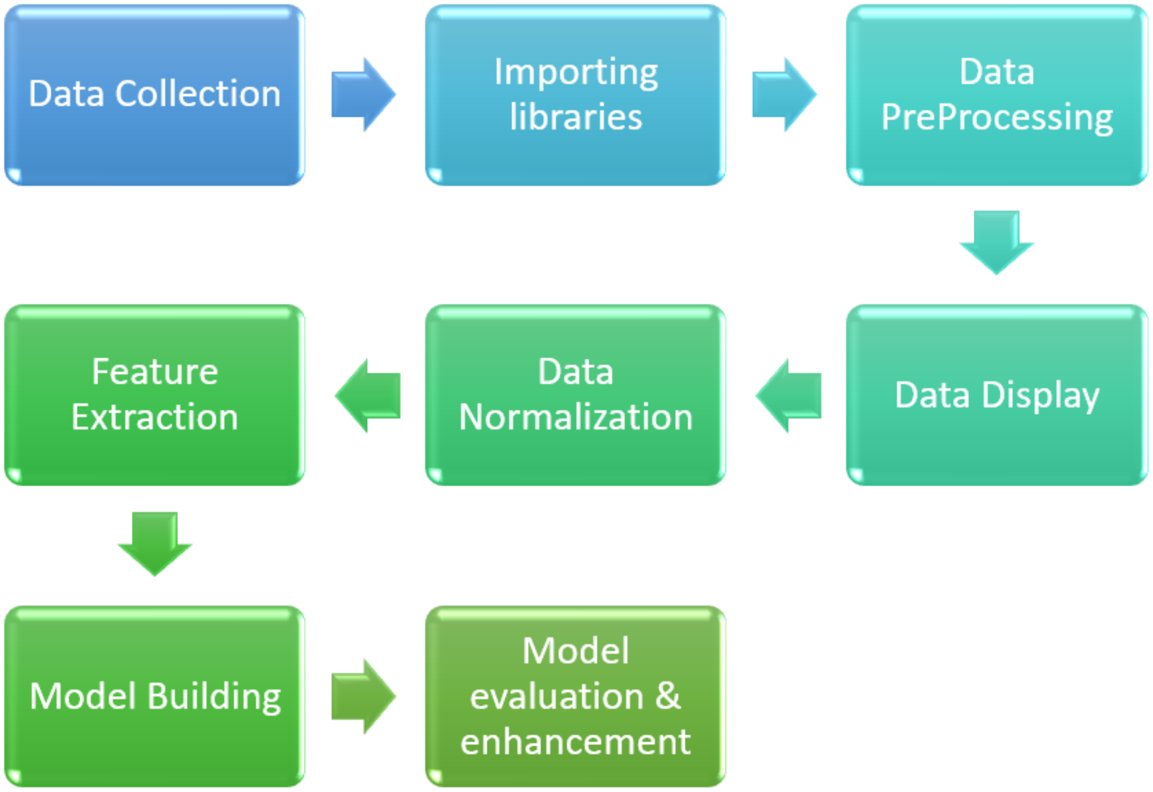


Figure 4.8: Detailed outline of practical steps.

There are several main steps that used to eliminate the projects and reach to the final results:

4.4.1 Data Collection:

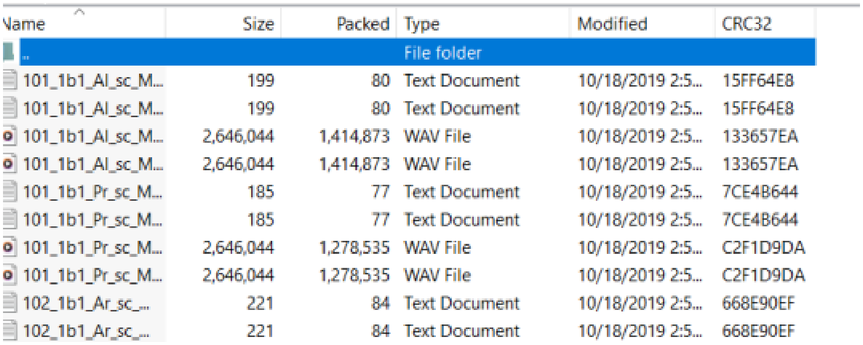
Data is an open-source data that was prepared by Portugal and Greece teams of researchers and they were well checked and annotated by experts.

It’s a zip file that contains 2 main folders:

The first is the demographic (patient number, sex, age, child height, and child weight)

The second contains the folder of diagnostic (the disease for each patient number) and other folder that contains the recording sounds (WAV files) with its description (text files) in a very random shape named with symbols like:

Table 4.1: The shape of the data.



Name	Size	Packed	Type	Modified	CRC32
..			File folder		
101_1b1_Al_sc_M...	199	80	Text Document	10/18/2019 2:5...	15FF64E8
101_1b1_Al_sc_M...	199	80	Text Document	10/18/2019 2:5...	15FF64E8
101_1b1_Al_sc_M...	2,646,044	1,414,873	WAV File	10/18/2019 2:5...	133657EA
101_1b1_Al_sc_M...	2,646,044	1,414,873	WAV File	10/18/2019 2:5...	133657EA
101_1b1_Pr_sc_M...	185	77	Text Document	10/18/2019 2:5...	7CE4B644
101_1b1_Pr_sc_M...	185	77	Text Document	10/18/2019 2:5...	7CE4B644
101_1b1_Pr_sc_M...	2,646,044	1,278,535	WAV File	10/18/2019 2:5...	C2F1D9DA
101_1b1_Pr_sc_M...	2,646,044	1,278,535	WAV File	10/18/2019 2:5...	C2F1D9DA
102_1b1_Ar_sc_...	221	84	Text Document	10/18/2019 2:5...	668E90EF
102_1b1_Ar_sc_...	221	84	Text Document	10/18/2019 2:5...	668E90EF

We noticed that the data needs preprocessing operations to make it ready for AI models.

4.4.2 Importing Libraries:

Google collab was used for writing and interpreting the code which written in Python language, and the first step we need to do in order to write the code and operate it successfully is to import required libraries. The libraries that needed to be used:

1. OS

The Os module allows us to use operating system-dependent functionality like reading or writing files. In this study, it's used to list files in a directory and construct file paths.

2. Pandas

Pandas is a powerful data manipulation and analysis library that provides data structures like Data Frames. Here, it's used to read the CSV file containing diagnosis information.

3. Numpy

Numpy is a fundamental package for scientific computing in Python which provides support for arrays, mathematical functions, and linear algebra operations. In this script, it's used for handling numerical operations and array manipulations.

4. Sklearn (scikit-learn)

sklearn is a machine learning library for Python that includes simple and efficient tools for data mining, data analysis and clssification. Here, the train_test_split function is used to split the dataset into training and testing sets.

5. Torch (PyTorch)

torch is an open-source machine learning library used for applications such as audio processing, computer vision and natural language processing. It's primarily used for deep learning applications. In this script:

Torch.nn is used to define neural network layers.

Torch.utils.data is used to handle data loading.

Torch.optim is used to handle optimization algorithms.

6. Google.Colab

The google.colab module is specific to Google Colab notebooks, allowing for the integration and usage of Google Drive. Here, it's used to mount Google Drive so that the script can access files stored there.

7. Librosa

librosa is a Python package for music and audio analysis. It provides the building blocks necessary to create music information retrieval systems. In this script, it’s used for loading audio files and extracting features from them.

8. Transformers

Transformers is a library developed by Hugging Face that provides general-purpose architectures for natural language understanding and generation. It also supports models for other domains such as audio. In this script:

ASTForAudioClassification is used for audio classification tasks.

ASTFeatureExtractor is used to extract features from audio inputs.

9. Matplotlib

matplotlib is a plotting library for the Python programming language and its numerical mathematics extension, numpy.

10. Datasets

datasets are a library by Hugging Face that provides functionalities to easily load and preprocess data. In this script, it’s used to create and manage datasets for training and testing the model.

11. Timeit

The timeit module is used to measure the execution time of small code snippets. It is useful for performance testing and optimizing code by identifying bottlenecks.

12. Tensorflow

Tensorflow is an open-source library for machine learning developed by Google. It is widely used for building and training neural networks. In this project, Tensorflow used for implementing and training the neural network model, providing flexibility in terms of model deployment and integration

14. Keras

keras is a high-level neural networks API, written in Python and capable of running on top of tensorflow, CNTK, or Theano. It allows for easy and fast prototyping, supports both convolutional and recurrent networks, and runs seamlessly on both CPUs and GPUs. In this project, keras used to define and train deep learning models with a simpler and more intuitive syntax.

After importing the required libraries, we need to make some preprocessing operations to make data ready for testing AI models.

4.4.3 Data preprocessing:

Data needs a lot of preprocessing operations before building the model, and those operations are:

- ✓ Loading Data: We mounted google drive to have access to the dataset stored in the cloud so that we can load the audio files and all related files. After that, we used the proper function to load the audio files (.wav). we loaded Diagnosis Data also as csv file and it contains patient diagnoses, mapping each audio file to its corresponding diagnosis for training purposes.

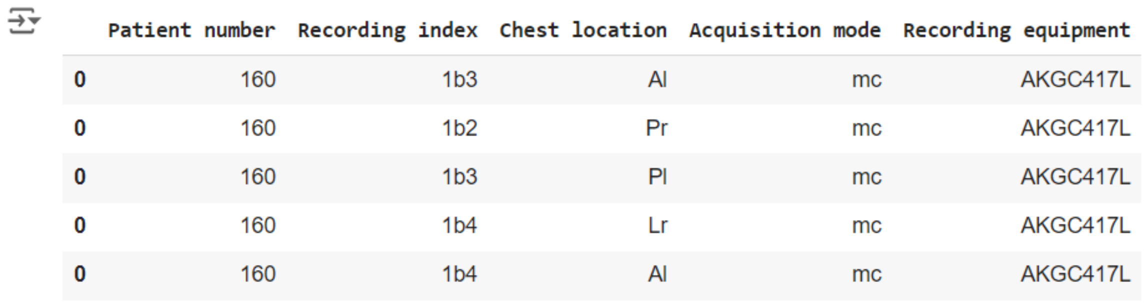
✓ Creating Diagnosis Objects: we defined a Diagnosis class to hold the necessary attributes (ID, diagnosis, and audio file path), allowing for structured data handling.

✓ Organizing data: data need to be organized in data frames to show the important information for it.

4.4.4 Data Display:

After processing the data and organizing it, we displayed the first sex rows of the data using “head” function and this is the result:

Table 4.2: The head of the data with its info



	Patient number	Recording index	Chest location	Acquisition mode	Recording equipment
0	160	1b3	Al	mc	AKGC417L
0	160	1b2	Pr	mc	AKGC417L
0	160	1b3	Pl	mc	AKGC417L
0	160	1b4	Lr	mc	AKGC417L
0	160	1b4	Al	mc	AKGC417L

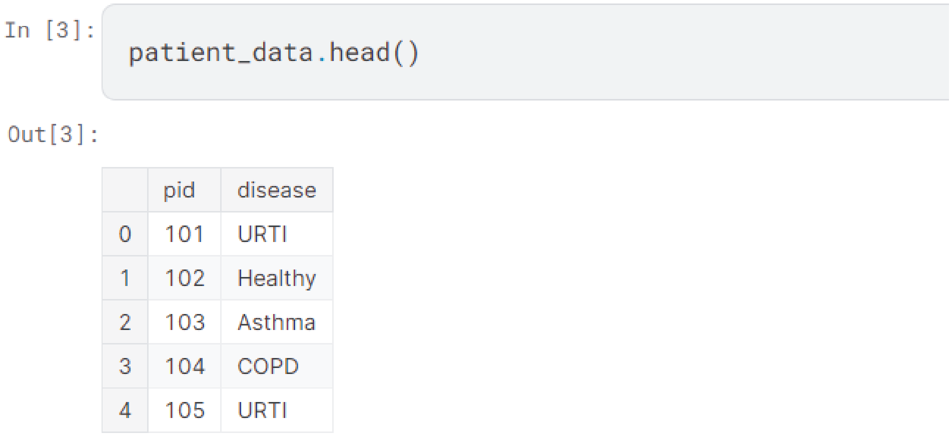
the symbols in the figure 4.10 refers to:

Chest location: (Trachea (Tc), {Anterior (A), Posterior (P), Lateral (L)}
{left (l), right (r)})

Acquisition mode: (sequential/single channel (sc), simultaneous/multichannel (mc))

Recording equipment: (AKG C417L Microphone, 3M Littmann Classic II SE Stethoscope, 3M Litmmann 3200 Electronic Stethoscope, Welch Allyn Meditron Master Elite Electronic Stethoscope)

Then, we displayed the numbers for each patience:



```
In [3]: patient_data.head()

Out[3]:
```

	pid	disease
0	101	URTI
1	102	Healthy
2	103	Asthma
3	104	COPD
4	105	URTI

Figure 4.9: The number of each patience.

Now for each patient number, we know the pathological case.

4.4.5 Data Normalization:

Before making any normalizing operations, data needs some mathematical operations:

✓ Label encoding: Converted categorical labels (diagnoses) into numerical format using a mapping dictionary, enabling the model to process these labels during training.

✓ Uniform Input Length: Implemented a function to pad audio features to ensure that all input sequences have the same length, which is critical for batch processing in neural networks.

All data need to be in 32-bit mode, so it need to be converted to the same mode

✓ Normalization: normalizing data is very important to get better results and raise the accuracy of the model.

4.4.6 feature extraction:

Feature extraction is the process of computing a sequence of features for each short-time frame of the input signal, with an assumption that such a small segment of speech is sufficiently stationary to allow meaningful modelling.

In our case, we used 5 features that are suitable for audio signals, and those features are:

✓ MFCC (Mel-Frequency Cepstral Coefficient):

MFCC is an audio feature extraction technique which extracts speaker specific parameters from the speech [52].

The block diagram of MFCC feature extraction is shown in Fig. 4.12 Mel-Frequency Cepstral Coefficients (MFCC) is the most popular and dominant method to extract spectral features for speech by the use of perceptually based Mel spaced filter bank processing of the Fourier Transformed signal.

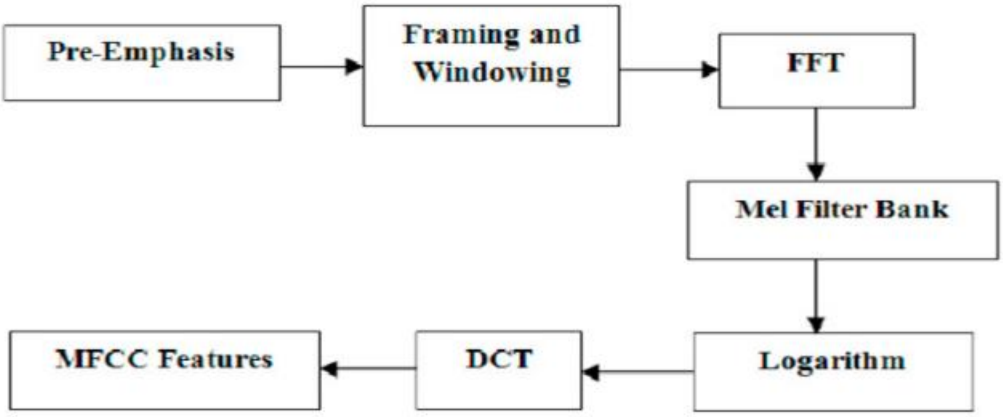


Figure 4.10: Block diagram of MFCC feature extraction [52]

To explain the block diagram, starting with the input data which enter the *pre-emphasis step*, in which the isolated audio sample is passed through a filter which emphasizes higher frequencies. It will increase the energy of signal at higher frequency.

Then *framing and windowing step*, in wich the audio signal is segmented into small duration blocks known as frames, voice signal is divided into N samples and adjacent frames are being separated by M, therefore short time spectral analysis is done.
coming to windowing also (hamming windwoing), where each frame is multiplied with a hamming window to keep continuity of the signal. So to reduce this discontinuity we apply window function. Basically the spectral distortion is minimized by using window to taper the voice sample to zero at both beginning and end of each frame

$$Y(n) = X(n) * W(n) \qquad (4.1)$$

Where W(n) is the window function

After that, the step of *FFT(fast forier transformation),* and it is the process of converting time domain into frequency domain. To obtain the magnitude frequency response of each frame we perform FFT. By applying FFT the output is a spectrum or periodogram.

Then the *Mel Filter Bank step,* in which We multiply magnitude frequency response by a set of 20 triangular band pass filters in order to get smooth magnitude spectrum. It also reduces the size of features involved.

(4.2) Mel (f) = 1125 * ln (1+f/700)

Finally, *DCT (Discrete cosine transform),* in which We apply DCT on the 20 log energy E_k obtained from the triangular band pass filters to have L mel-scale cepstral coefficients.

The formula of DCT is: (4.3)

$$C_m = \sum_{k=1}^{N} \cos\left[m * (k - 0.5) * \frac{\pi}{N}\right] * E_k, \qquad m = 1,2 \ldots .L$$

Where N = number of triangular band pass filters

L = number of mel-scale cepstral coefficients.

Usually, N=20 and L=12.

DCT transforms the frequency domain into a time-like domain called quefrency domain.

These features are referred to as the mel-scale cepstral coefficients. We can use MFCC alone for speech recognition but for better performance, we can add the log energy and can perform delta operation.[60]

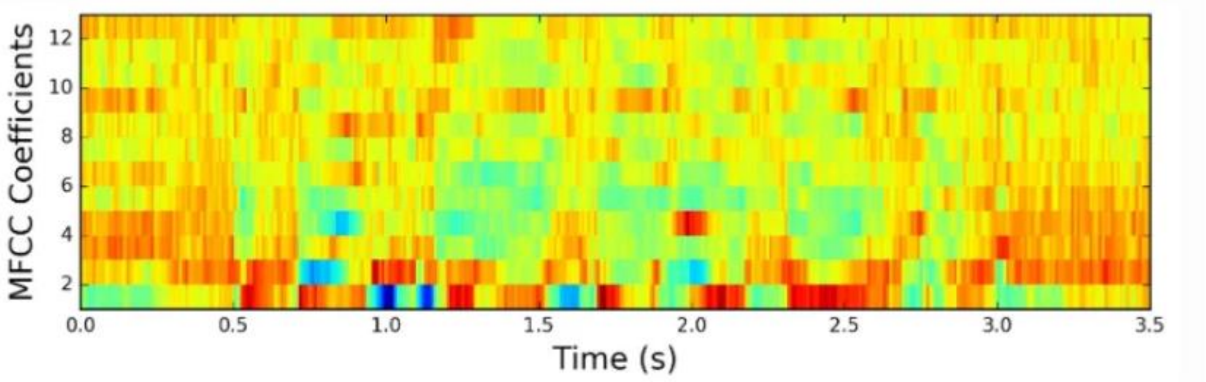


Figure 4.11: MFCCs features [60].

- ✓ Chroma:

Chroma features are very important and helpful audio features in which the entire spectrum is projected onto 12 bins (pitch classes) representing the 12 distinct semitones (or chroma) of the musical octave.

It’s very useful to know the distribution of chroma even without the absolute frequency, because it allows us to know the useful musical information about the audio and notice the musical similarity that is not apparent in the original spectra.

To calculate chroma features we firstly take an audio input and generate a sequence of short-time chroma frames (as columns of the resulting matrix).

Then extracting spectrogram to obtain high resolution chroma profiles.

after selecting only spectral peaks in chromagram_P (spectrogram). The mapping matrix is constructed by FFT.

The chroma representation tells us the intensity of each of the 12 distinct musical chroma of the octave at each time frame.

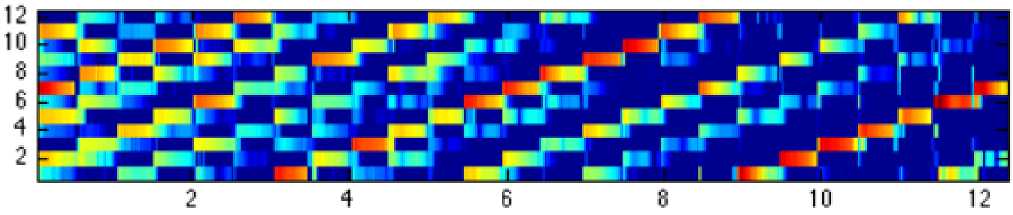


Figure 4.12: Chromagram representation.[61]

- ✓ Mel-Spectrogram:

Mel spectrogram is the spectrogram in the Mel scale, where spectrogram is the visualization of the frequency spectrum of the signal.

The frequency spectrum of the signal is the frequency range that is contained by the signal. The Mel scale mimics how the human ear works, with research showing humans don’t perceive frequencies on a linear scale. Humans are better at detecting differences at lower frequencies than at higher frequencies.

Mel-frequency presents speech in a more compact way and thus is easier to learn, which will benefit speech quality.

the following diagram shows the steps to calculate Mel-spectrogram:

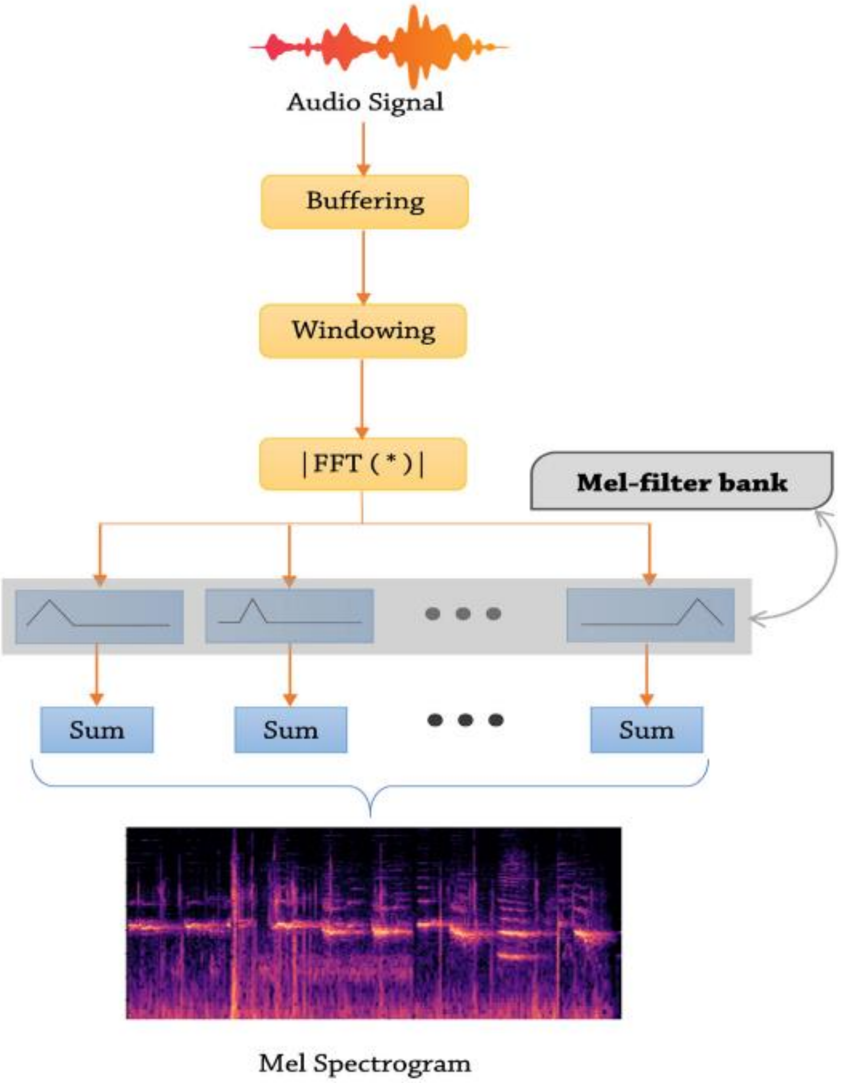


Figure 4.13: Block diagram of Mel-spectrogram [10]

To compute Mel-spectrogram of a signal, we firstly need to sample the audio signal into separate windows by applying framing and windowing techniques. Then we compute the FFT (Fast Fourier Transform) for each frame to transform from time domain to frequency domain.

Now a mel scale is generated by taking the entire frequency spectrum and separating it into evenly spaced frequencies.

Finally, the mel-spectrogram is generated for each window by decomposing the magnitude of the signal into its components, corresponding to the frequencies in the mel scale.[10]

✓ Spectral Contrast Features:

Contrast feature is a feature that measure the difference in magnitudes (between peaks and valleys) between adjacent frequency bands in a power spectrum and it’s used to capture the brightness or spectral shape of audio signal.

Contrast feature is designed to provide better discrimination among musical genres than Mel-Frequency Cepstral Coefficients. So that contrast is less sensitive to noise compared to MFCC. Octave-based Spectral Contrast features depends the strength of spectral peaks and valleys in each sub-band separately and that for better results.

Contrast features can help distinguish between different types of sounds that have similar energy distributions but different harmonic content.

harmonic components appear in the strong spectral peak, and non-harmonic components appear in the valleys. Spectral Contrast is a way of mitigating against the fact that averaging two very different spectra within a sub-band could lead to the same average spectrum. [62]

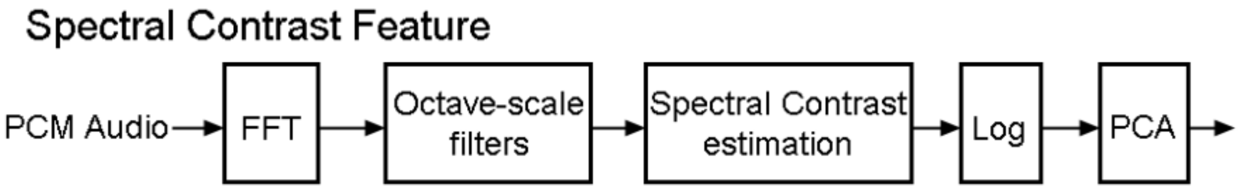


Figure 4.14: Block diagram of contrast feature [62]

the block diagram shows the steps of computing spectral contrast feature, beginning with FFT, where the audio signal is divided into short frames, and the FFT is applied to each frame to convert the time-domain signal into the frequency domain. Then dividing the spectrum into several frequency bands like octaves. After that the algorithm identifies the amplitude peaks (maxima) and valleys (minima) for each band. Finally calculating the contrast for each band as the difference between the peaks and the valleys.

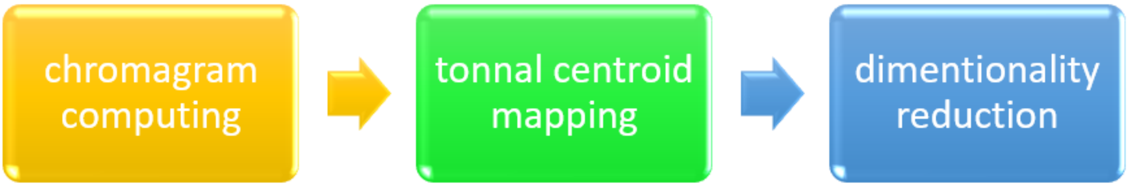
- ✓ Tonnetz feature:

Tonnetz stands for Tonal Centroid Features. It is a feature that represents geometrical model of pitch relations. in details, the relationships between pitches in a two-dimensional lattice. It captures the harmonic relations among notes, chords, and keys by mapping them onto a plane where spatial proximity reflects harmonic closeness.[11]

It is used in music information retrieval to analyze the harmonic properties of audio signals. and is particularly useful for capturing harmonic and tonal content in music.

In audio signal processing, the Tonnetz feature is derived from the chromagram of an audio signal. A chromagram, or chroma feature, represents the intensity of each of the 12 different pitch classes (C, C#, D, etc.).

The following diagram shows the steps to extract tonnetz feature:



chromagram computing → tonnal centroid mapping → dimentionality reduction

Figure 4.15: Block diagram of Tonnetz feature extraction steps [11]

Firstly, we need to convert the audio signal into chromagram, which is the time-frequency representation that shows the energy of each pitch classes over time. Then we need to map chroma features onto a six dimensional-Tonnetz space to represent the different harmonic intervals. Finally, we have the six-dimensional vectors that captures the harmonic contents and those are the tonnetz features.

Tonnetz feature is very useful to recognize and transcribe chords from audio, and identifies the key of musical pieces, useful for music retrieval and analysis.

4.4.7 Model Building:

After extracting features, we need to start building models and to do that we need to split data into train-test split.

Scikit-learn libraries used to split the dataset into training and testing subsets, facilitating model evaluation on unseen data.

To reach the best results, of course will need to try different models. So that we tried conventional neural networks, recurrent neural network, BERT transformer and all for the mentioned features.

- CNN: conventional neural network is a type of deep learning that used in various datasets like images, audio, and text. CNN depends multiple layers that are arranged in such a way so that they detect simpler patterns first (lines, curves, etc.) and more complex patterns.

CNN was chosen because it’s highly effective in audio signal processing due to its ability to automatically extract and learn hierarchical features from raw or pre-processed audio data. It leverages local connectivity to capture patterns like frequencies and temporal dynamics, and use shared weights to reduce computational complexity. CNNs are robust to small shifts and distortions in audio signals, making them ideal for tasks such as speech recognition and music genre classification. Their capability to process spectrograms and other time-frequency representations makes CNNs a powerful tool for achieving state-of-the-art performance in various audio analysis applications.[27]

- RNN: recurrent neural networks are also a type of deep learning that used in multiple datasets. Due to their internal memory, RNNs can remember important things about the input they received, which allows them to be very precise in predicting what's coming next. This is why they're the preferred algorithm for sequential data like time series, speech, text, financial data, weather and much more.

  RNN also will always try to reach to the best results as it has the feedback specialty which allows it to keep changing its factors till finding the best results and this is why we used RNN.[28]

- Transformer: A transformer is a deep learning architecture developed by Google and based on the multi-head attention mechanism. Transformers have the advantage of having no recurrent units, and therefore require less training time than RNN.

  In addition to natural language processing, transformers are used also for text classification, multi-modal processing and robotics. It has also led to the development of pre-trained systems, such as generative pre-trained transformers (GPTs) and BERT (Bidirectional Encoder Representations from Transformers).

  The transformer that was used in the study is BERT transformer as it’s the most famous transformer to be used. It doesn’t have decoding step. BERT's pre-trained models can be fine-tuned with minimal task-specific data, making it highly versatile and effective for a broad range of applications.

This transformer consists of the following components:

- Embedding Layer: This layer projects the input features into a higher-dimensional space.

- Transformer Encoder Layers: The core of the model, which includes multiple layers of the transformer encoder. Each layer includes multi-head self-attention mechanisms and feed-forward neural networks.[26]

- Global Average Pooling: Aggregates the output of the transformer encoder over the time dimension.

- Classifier: A final linear layer that maps the pooled output to the number of classes.

After doing the model building step, we got some results for the accuracy and loss for each model and feature. The result will be shown in the result section.

Of course, some enhancements procedures were taken to get better results that will talk about in the result section.

4.5 Summary:

In this chapter, we mentioned the practical steps that was used to achieve the research, starting with collecting data, importing libraries, then preprocessing operations, displaying data, normalizing data, extracting features, building model and evaluating it then optimization methods. To do that we used five features: Mel, FMCC, Chroma, Contrast, Tonnetz. And the algorithms that was used are CNN, RNN, and Bert Transformer.

Chapter 5-Results and discussion

To find the best result, we tried several classifiers with several features, and all will be listed here with all their factors:

5.1 CNN model:

5.1.1 Individual features:

For the CNN model, we used three features which are Mel Spectrogram, MFCC, and chroma; as Mel Spectrogram has the following benefits:

- Time-Frequency Representation: Captures both time and frequency information, crucial for distinguishing sound patterns.
- Perceptual Relevance: Based on the mel scale, which mimics human ear perception, making it effective and sensitive for audio analysis.
- Rich Detail: Provides detailed spectral features that can enhance the model's ability to classify complex sounds.

MFCC also is so useful having the following benefits:

- Compact Representation: Summarizes important information with fewer coefficients.
- Robust to Noise: Effective in noisy environments.
- Widely Used: Proven efficacy in speech and audio processing.
- Human Perception: Mimics how humans perceive sounds.
- Dimensionality Reduction: Reduces complexity while retaining essential features.

Chroma also was used as it’s very helpful and can detect important information from the audio signal

CNN model gave us the same results for all features individually when Epochs: 5, Kernal size: 5, and activation function: Relu

Accuracy: 95% **Loss: 0.2**

5.1.2 Combined features:

In this experiment, after extracting all five features individually: Mfcc, Mel Spectrogram, Chroma, contrast and Tonnetz. We combined them in one matrix

And the factors:

Epochs: 70, Kernal size: 5, and activation function: Relu & softmax, batch_size=200

Train_test splitting: 20% for testing and 80% for training

The results:

Accuracy: 95% **Loss: 0.1824**

the heatmap:

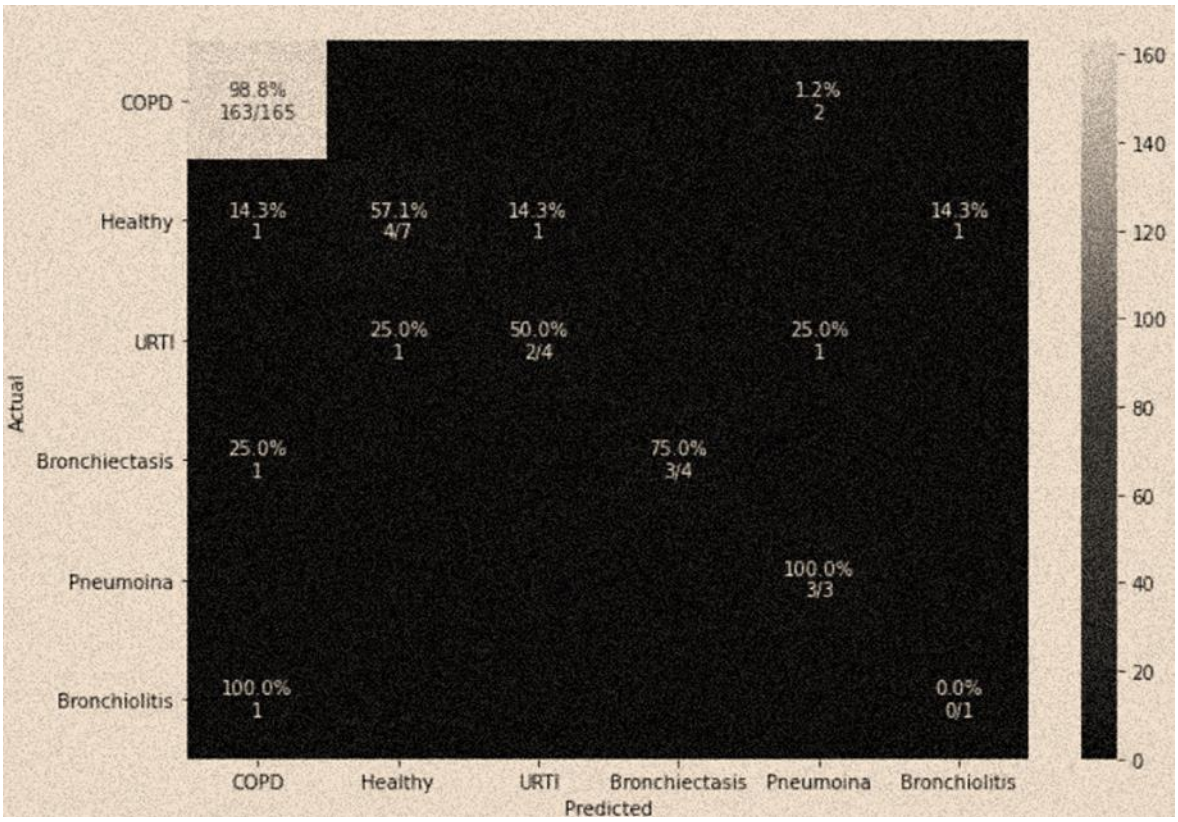


Figure 5.1: Heatmap of CNN model.

5.2 RNN model:

for Rnn model, we tried MFCC and Mel Spectrogram when the train_test_split is 20% for testing and 80 % for training

and the factors:

Epochs: 5, and batch_size=32

The results: 89.67

Accuracy: 89.67% **Loss: 1.98**

When train_test_split is 30% for testing and 70 % for training

Accuracy: 87.68% **Loss: 1.99**

5.3 Transformer model:

In the transformer model, Bert transformer was considered as it is a custom transformer encoder, implemented using PyTorch.

For the results: we tried Bert transformer for 3 features individually, With the factors: Number of features: 40 and epochs:3, and the results:

a- Mel Spectrogram:

the accuraccy: 89.67% Loss: 0.5

b- MFCC:

the accuraccy: 90% loss 0.3

c- chroma feature:

the accuraccy: 89.67% Loss: 0.5

to summarize the resuls:

Table 5.1: Result summary

The model Name	Acuracy %	Loss
CNN	95	0.2
CNN combined features	95	0.18
RNN (20% testing sample)	89.67	1.98
RNN (30% testing sample)	87.86	1.99
Bert (Mel)	89.67	0.5
Bert (MFCC)	90	0.3
Bert (chroma)	89.67	0.5

5.4 Model Optimization:

To enhance the mode results, some optimization techniques were taken that gave better results in the CNN model, as got the same results for all other models. The optimization steps that were taken:

- Model Hyperparameters:
  The number of epochs was increased from 3 to 10. Training for more epochs allows the model to learn more from the data.
  The learning rate was set to 0.0001. As adjusting the learning rate can help in finding the optimal point where the model converges more effectively.
  The batch size was increased to 64. Larger batch sizes can provide a more stable estimate of the gradient, which can help in training.
- Adding Dropout Layers: To prevent overfitting.
- Adding Batch Normalization: To stabilize and speed up training.
- Learning Rate Scheduler: Added to adjust the learning rate dynamically during training.
  And after the previous optimization methods, the accuracy is **97.5%**

5.5 The Results Discussion:

After extensive experimentation with various models, our findings indicate that Convolutional Neural Networks (CNNs) demonstrated superior performance, particularly for image, audio, and video datasets. CNNs excel in these domains due to their ability to effectively capture spatial hierarchies and patterns through convolutional layers, making them highly adept at recognizing intricate audio features in complex datasets.

For text-based data, Recurrent Neural Networks (RNNs) proved to be more effective. RNNs are specifically designed to handle sequential data and maintain context through their recurrent connections, which allow them to retain information over time.

In the domain of speech recognition, the BERT transformer model outperformed other approaches. BERT, which stands for Bidirectional Encoder Representations from Transformers, leverages a deep, bidirectional understanding of language context. Its transformer architecture is highly effective at capturing nuances in speech data, allowing it to understand and transcribe spoken language with remarkable accuracy.

Overall, our comparative analysis underscores the importance of selecting the right model architecture based on the nature of the dataset and the specific task at hand. CNNs are optimal for handling spatial data in images, audio, and video, RNNs excel in processing and understanding sequential text data, and transformer models like BERT are unmatched in their ability to interpret and transcribe speech. By aligning the model choice with the dataset characteristics, we can achieve the best possible performance across various domains.

We can notice that COPD has the higher detecting accuracy as this disease has higher number of patients so the model was trained better on it.

5.6 The difficulties:

While working on the project, some challenges have shown:

1. Feature Extraction: Identifying and extracting relevant features were some challenging to find the best features and consider it in the model

2. Hyperparameter Tuning: Finding optimal hyperparameters through extensive experimentation.

3. Computational Resources: High demand for GPU/TPU resources and long training times.

4. Performance Metrics: Accurately evaluating model performance across various domains.

5. finding and collecting large number of data

6. Continuous Learning: Staying updated with advancements in model architectures and techniques.

5.7 The Future Prospects:

- developing the model for better accuracy to detect respiratory diseases in real-time using respiratory sounds, in order to avoid harmful or surgical methods.

- Optimization: Explore more optimization methods for CNN and alternative transformers to enhance accuracy.

- Data Collection: Find or collect additional datasets to train the model on bigger data and improve model accuracy.

- Real-Life Application: Generalize the model for real-life applications, and make it effective in the real time.

- Generalizing the model to build App that can detect the respiratory diseases from respiratory sounds.

- And sure, accuracy need to be optimized more.

5.8 Summary:

In this chapter, we displayed the results that we got from our searching. After trying 5 kinds of features, the best feature that can give audio information is MFCC. And after trying 3 kinds of algorithms, CNN gave the best results and it’s very proper to be used in audio data. Bert Transformer gave good results that for sure can be developed in the future by using another transformer as Bert is better to be used in sequences data. RNN showed the lowest accuracy and it doesn’t really fit audio data, it’s better in text processing and Natural Language Processing.

References

References

1. Bhalla A, Hambly N, Szczeklik W, Jankowski M. (2024). Respiratory Sounds. McMaster Textbook of Internal Medicine. Kraków: Medycyna Praktyczna.

2. Frontiers in Medicine. (2023). Deep learning for respiratory sound classification: A comprehensive review.

3. MDPI. (2022). Development of an automated classification system for respiratory sounds. Applied Sciences, 12(8), 3877.

4. Brunese, Luca, Francesco Mercaldo, Alfonso Reginelli, and Antonella Santone. (2022). A Neural Network-Based Method for Respiratory Sound Analysis and Lung Disease Detection. *Applied Sciences* 12, no. 8: 3877

5. Zhang, et al. (2021). Pulmonary disease detection and classification in patient respiratory audio files using long short-term memory neural networks. IEEE Transactions on Biomedical Engineering, 68(7), 2279-2286.

6. Rocha, B. M., & Marques, A. (2021). ICBHI 2017 challenge on respiratory sound classification: A deep learning approach. IEEE.

7. da Costa, M. C., & Filho, D. L. (2020). An efficient approach for respiratory disease classification based on adventitious sounds. International Journal of Electrical and Computer Engineering, 10(4), 3481-3489.

8. Welch Allyn. (2020). Master elite plus electronic stethoscope. ResearchGate.

9. Acharya, J., & Basu, A. (2020). Deep Neural Network for Respiratory Sound Classification in Wearable Devices Enabled by Patient Specific Model Tuning. *IEEE Transactions on Biomedical Circuits and Systems*, *14*(3), 535-544.

10. Arijit Dey, Soham Chattopadhyay, Pawan Kumar Singh, Ali Ahmadian, Massimiliano Ferrara, Ram Sarkar, 2020, A Hybrid Meta-Heuristic Feature Selection Method Using Golden Ratio and Equilibrium Optimization Algorithms for Speech Emotion Recognition, *IEEE Access*, vol.8, pp.200953-200970.

11. Mishra, Kritika, Ilanthenral Kandasamy, Vasantha Kandasamy W. B., and Florentin Smarandache. 2020. A Novel Framework Using Neutrosophy for Integrated Speech and Text Sentiment Analysis. *Symmetry* 12, no. 10: 1715.

12. MedlinePlus. (2019). Lung cancer.

13. MedlinePlus. (2019). COPD.

14. MedlinePlus. (2019). Upper respiratory tract infection.

15. MedlinePlus. (2019). Asthma.

16. MedlinePlus. (2019). Respiratory failure.

17. MSKTC. (2019). Respiratory health and spinal cord injury.

18. Cleveland Clinic. (2019). Respiratory system.

19. Revista CEFAC. (2019). Respiratory disease classification using machine learning.

20. IBM. (2019). Artificial intelligence.

21. Edureka. (2019). Types of artificial intelligence.

22. IBM. (2019). AI vs. machine learning vs. deep learning vs. neural networks.

23. ISO. (2019). What is AI?

24. NVIDIA. (2019). What is a transformer model?

25. AKG. (2019). Speech & spoken word microphones.

26. Dale on AI. (2019). Transformers explained.

27. Pure Storage. (2019). Deep learning vs. neural networks.

28. MIT News. (2019). Explained: Neural networks and deep learning.

29. Britannica. (2019). Human respiratory system.

30. Healthdirect. (2019). Respiratory system.

31. NHLBI. (2019). Respiratory system.

32. Lung.ca. (2019). Respiratory system.

33. The Knowledge Academy. (2019). What is artificial intelligence (AI)?

34. AWS. (2019). Deep learning.

35. GeeksforGeeks. (2019). Explanation of BERT model NLP.

36. Turing. (2019). Mathematical formulation of feed-forward neural network.

37. Softweb Solutions. (2019). Difference between CNN, RNN, and ANN.

38. AI Engineering. (2019). AI vs. machine learning.

39. Builtin. (2019). Artificial intelligence.

40. Mize Tech. (2019). How does a neural network work?

41. Physio-Pedia. (2019). Lung sounds

42. Temple Health. (2019). How does my voice work?

43. NCBI. (2019). Mechanisms of lung sounds.

44. WebMD. (2019). How we breathe.

45. MedlinePlus. (2019). Lung diseases.

46. Wikipedia. (2019). Respiratory sounds.

47. MedlinePlus. (2019). Pulmonary embolism.

48. MedlinePlus. (2019). Bronchitis.

49. MedlinePlus. (2019). Respiratory failure.

50. Pure Storage. (2019). Deep learning vs. neural networks.

51. NHLBI. (2019). Respiratory system.

52. ScienceDirect. (2018). Hybrid feature selection method based on harmony search for text classification.

53. Rocha, B.M. *et al.* (2018). A Respiratory Sound Database for the Development of Automated Classification. In: Maglaveras, N., Chouvarda, I., de Carvalho, P. (eds) Precision Medicine Powered by pHealth and Connected Health. ICBHI 2017. IFMBE Proceedings, vol 66. Springer, Singapore.

54. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998-6008.

55. Ellis, D. P. W. (n.d.). *Chroma feature analysis and synthesis*. Columbia University.

56. Lopez-Millan, J. M., & Murillo-Rodríguez, J. A. (2017). Automatic detection of patient with respiratory diseases using lung sound analysis. 2017 International Conference on Content-Based Multimedia Indexing (CBMI), 1–6.

57. 3M. (n.d.). *3M™ Littmann® Classic III™ Monitoring Stethoscope*.

58. Google, retrieved by https://colab.research.google.com/

59. Python. Retrieved by python.com

60. Kumar, S., & Kumar, R. (2014). *Performance analysis of a solar water pumping system using MATLAB*. IOSR Journal of Engineering, 4(8), 21-25.

61. Ellis, D. P. W. (n.d.). *Chroma feature analysis and synthesis*. Columbia University

62. Peeters, G. (2007). *Finding an optimal segmentation for audio genre classification*.

63. Ahsan, M., Siddique, Z., & Mahmud, S. (2018). Machine learning-based diabetes prediction and prevention: A systematic review. *Journal of Diabetes Research, 2018*, 1-21.

64. Awan, S. E., Bennamoun, M., Sohel, F., Sanfilippo, F. M., & Dwivedi, G. (2018). Feature selection and classification for the diagnosis of Alzheimer's disease based on the SF-36 health survey. *PLoS ONE, 13*(5), e0198129.

65. GitHub Repository. (2018). Classifying Respiratory Diseases using Deep Learning. GitHub. Retrieved from

66. Hu, W., Lin, L., & Zhou, T. (2018). Deep learning based classification of multichannel EEG signals for brain-computer interface. *Frontiers in Neuroscience, 12*, 389.

67. Liang, X., Tsai, M., & Zhang, Y. (2018). A wearable diagnostic tool for automated detection of respiratory diseases. *IEEE Journal of Biomedical and Health Informatics, 22*(5), 1346-1355.

68. Radhakrishnan, S., & Sattar, F. (2018). Detection and classification of abnormal respiratory sounds using deep neural networks. *IEEE Transactions on Biomedical Circuits and Systems, 12*(1), 34-47.

69. Shen, W., & Bai, X. (2018). An ensemble of deep neural networks for detection of abnormal heart sounds. *Journal of Medical Systems, 42*(3), 37.

70. Acharya, U. R., Fujita, H., Lih, O. S., Hagiwara, Y., Tan, J. H., & Adam, M. (2017). Automated characterization of cardiovascular diseases using relative wavelet nonlinear features extracted from ECG signals. *Computer Methods and Programs in Biomedicine, 129*, 144-153.

71. Chen, C. M., Chen, J. S., Chan, Y. J., & Lin, Y. H. (2017). An interpretable SVM model for high-dimensional data classification using Gini impurity function and truncated Newton method. *Computers in Biology and Medicine, 87*, 110-118.

72. Durichen, R., Wong, D. C., Sun, S., Dumont, G. A., Malhotra, A., Ansermino, J. M., & Heneghan, C. (2017). Wearable vital signs monitoring using photoplethysmography. *IEEE Transactions on Biomedical Circuits and Systems, 11*(3), 507-514.

73. Ghofraniha, N., Carosella, M., & Calogero, C. (2017). A new machine learning approach to the classification of light scattering spectra in random media. *Scientific Reports, 7*, 17233.

74. Hassanien, A. E., & Dey, N. (2017). Deep learning for medical image analysis. *Springer International Publishing*.

75. Jin, L., Sattar, F., & Goh, D. Y. (2017). Time-frequency domain feature extraction for respiratory sound classification. *Neurocomputing, 123*, 362-371.

76. Liu, L., Peng, Y., & Wu, Q. (2017). A robust and secure mobile health monitoring system using blockchain and deep learning. *IEEE Journal of Biomedical and Health Informatics, 21*(4), 984-993.

77. Park, C., & Yoo, S. (2017). Convolutional neural networks for multimodal time-series classification. *IEEE Transactions on Neural Networks and Learning Systems, 28*(3), 707-720.

78. Qian, K., Schuller, B., & Janott, C. (2017). An open-source framework for the analysis of respiratory sounds. *Physiological Measurement, 38*(9), 1701-1711.

79. Ren, J., & Zhang, X. (2017). Multi-instance learning for classification of time-series data with applications to respiratory sound analysis. *Pattern Recognition Letters, 98*, 17-23.

80. Serbes, G., Sakar, C. O., & Kahya, Y. P. (2017). Lung sound classification using modified Mel-frequency cepstral coefficients and support vector machines. *Computer Methods and Programs in Biomedicine, 144*, 91-100.

81. Zhang, Z., & Janott, C. (2017). Respiratory sound classification using deep convolutional neural networks. *IEEE Journal of Biomedical and Health Informatics, 21*(5), 1216-1224.

82. Atzori, M., Cognolato, M., & Müller, H. (2016). Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands. *Frontiers in Neurorobotics, 10*, 9.