REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Information Technologies

A MACHINE LEARNING APPROACH TO DIFFERENTIATE BETWEEN ACUTE ASTHMA AND BRONCHITIS IN PRESCHOOL CHILDREN

Waleed Hameed Salih SALIH

Master's Thesis

Supervisor

Asst. Prof. Dr. Hakan KOYUNCU

Istanbul, 2022

A MACHINE LEARNING APPROACH TO DIFFERENTIATE BETWEEN ACUTE ASTHMA AND BRONCHITIS IN PRESCHOOL CHILDREN

Waleed Hameed Salıh SALIH

Information Technologies

Master's Thesis

ALTINBAŞ UNIVERSITY

2022

The thesis titled A MACHINE LEARNING APPROACH TO DIFFERENTIATE BETWEEN ACUTE ASTHMA AND BRONCHITIS IN PRESCHOOL CHILDREN prepared by WALEED HAMEED SALIH SALIH and submitted on 6/12/2022 has been **accepted unanimously** for the degree of Master of Science in Information Technologies.

Asst. Prof. Dr. Hakan KOYUNCU

the Supervisor

Thesis Defence Committee Members:

Asst. Prof. Dr. Hakan KOYUNCU	Department of Computer Engineering, Altinbaş University	__________________
Asst. Prof. Dr. Abdullahi Abdu IBRAHIM	Department of Computer Engineering, Altinbaş University	__________________
Asst. Prof. Dr. Hakan AYDIN	Department of Computer Engineering Topkapi University	__________________

I hereby declare that this thesis meets all format and submission requirements of a Master's thesis.

Submission date of the thesis to Institute of Graduate Studies: ___/___/___

I hereby declare that all information in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Waleed Hameed Salih SALIH

Signature

DEDICATION

At the beginning of my talk, I would like to thank God Almighty, who guided me and paved the way for me to complete my scientific thesis , I also extend my thanks and gratitude to my dear father, may God have mercy on him, who has been supportive of me throughout his life, my generous mother which she prays to God to help me in my studies, and my beloved wife who supported and encouraged me throughout the study period ,I also thank my children, Rawan, Rahaf, Hoor and Ibrahim, who are the flower of my life, and I pray to God to protect and guide them.

ACKNOWLEDGMENT

I extend my sincere thanks and gratitude to my supervisor assistant professor, Dr. Hakan KOYUNCU, who was in contact with me throughout the period of writing the thesis, answering my questions and guiding me to choose modern methods, as I thanks my brother, the consultant, Dr. Majeed Al-Ajeli, who helped me a lot in choosing the subject of the thesis and collecting the data that was examined by him inside Fallujah Teaching Hospital for Women and Children during a period of four months. We ask Allah to grant success.

ABSTRACT

A MACHINE LEARNING APPROACH TO DIFFERENTIATE BETWEEN ACUTE ASTHMA AND BRONCHITIS IN PRESCHOOL CHILDREN

Salih, Waleed Hameed Salih

M.Sc., Information Technologies, Altınbaş University

Supervisor: Asst. Prof. Dr. Hakan KOYUNCU

Date: 12/2022

Pages: 94

Among the common diseases that affects the lower respiratory tract infections (LRTI) of children around the world are acute asthma and bronchitis, it often occurs in preschool age and there are 12 clinical features overlapping between the two diseases, the most common of which are coughing, wheezing, runny nose, and shortness of breath. therefore, most people do not distinguish between two diseases, it is necessary to visit a specialist doctor to diagnose the cases and to receive the appropriate treatment for each case , due of the large number of cases that increase during weather fluctuations, such as high and low temperatures and environmental pollution such as fumes smoke and dust, etc., which need an accurate diagnosis, many junior doctors working in emergency halls face difficulties in diagnosing cases and the differentiation between the two diseases, so these doctors' resort to the consulting doctor to obtain a diagnosis that differentiates between the two diseases. In this study, we presented 3 machine learning models (K - NN , Decision Tree , MLP) and 2 deep learning models (1D-CNN, LSTM), where we trained this models on a text dataset consisting of 512 real cases that collected by the consultant paediatrician at Fallujah Teaching Hospital for Women and Children in Iraq during four months started in march 2022 to June 2022 and after using all modern methods the final results showed that the 1D-CNN outperformed the rest of the models with an accuracy of (99.3506) and ROC – AUC (99.32) which was chosen as a binary classifier to this study .

Keywords: Machine Learning, Deep Learning, Acute Asthma, Acute Bronchitis

TABLE OF CONTENTS

Pages

ABBREVIATIONS

ML	:	Machine Learning
DL	:	Deep Learning
KNN	:	K-Nearest Neighbours
DT	:	Decision Tree
MLP	:	Multi-Layer Perceptron
SLP	:	Single Layer Perceptron
CNN	:	Convolutional Neural Network
AUC	:	Area Under the Curve
LSTM	:	Long Short -Term Memory
COPD	:	Chronic Obstructive Pulmonary Disease
ROC	:	Receiver Operating characteristic Curve
FNN	:	Feedforward Neural Network
CV	:	Cross Validation
1D	:	One Dimensional
ED	:	Emergency Department
+ve	:	Positive
-ve	:	Negative
SOB	:	Shortness of Breath
LRTI	:	Lower Respiratory Track Infections

1. INTRODUCTION

1.1 BACKGROUND

Acute asthma and bronchitis are common diseases that affects the Lower Respiratory Tract Infections (LRTI) of children around the world and cause inflammation, airway irritation and coughing, these infections can be caused by a viral or bacterial infection , it is considered a main reason for the increase the number of deaths among children because the rapid spread of infection especially in the low income countries , these diseases can occur at any age, but they often start in preschool children (less than 6 years) , sometimes people confuse between acute asthma and bronchitis due there are overlap symptoms between the two diseases as cough and (SOB) , runny nose , wheeze , and makes it difficult to differentiate between them therefore, the junior doctors face difficult to differentiate between them [1],[2]. Acute asthma occurs for the first time in preschool age or cases diagnosed with previously asthma so that it has an acute or semi-acute onset , acute asthma can be caused by a viral infection or some allergens as result the air irritants as fumes and dust mites or smoke and strong odors such as perfumes or the causes may be include the deterioration of the health case of children as influenza, sinusitis, or upper respiratory infection, and in some cases because to family history[3].Symptoms of acute asthma different according to the child's immunity and age , often start with SOB , coughing , wheezing, and eczema ,these symptoms occur as result of contraction of the muscles surrounding the airways and the mucus production increases, making breathing difficult because of obstruction the airways , the duration and severity of the disease may differ in children, so some cases it lasts for a few minutes if the attacks are mild or lasts from hours to days if the attacks are severe therefore, it is necessary to visit the emergency department in the hospital due to the risk that threatens the lives of children, where doctors can identify of acute asthma exacerbation based on the examination and medical history of the patient , also acute asthma can be known as exacerbation of asthma[4]. Acute bronchitis is a self-limited viral infection of the upper airway, where cough appears, which is one of the first diagnostic symptoms of the disease[5]. Acute bronchitis is diagnosed after excluding other respiratory diseases as acute asthma and pneumonia and is the most common clinical illness in USA, acute bronchitis is caused by infection of the large airway and is usually due to viruses and bacterial infections uncommon , sometimes irritants bacteria and allergens such as polluted air, smoke and dust are one of the reasons for the emergence of bronchitis, it is estimated that (5%)

of the general population infects acute bronchitis each year during the flu season which usually occur in the winter and autumn, also sometimes acute bronchitis can occur as a result of the of infection of the upper respiratory tract, symptoms of the disease include a productive cough in addition to (SOB), wheezing, with a low grade fever, headache, runny nose and others usually, the cough continues after acute bronchitis for a period of 10-20 days and sometimes lasts for four weeks and , where the average cough is 18 days after acute bronchitis , sometimes acute asthma is diagnosed as acute bronchitis for one - third of patients who have acute cough (misdiagnosed)[6]. Acute asthma and bronchitis have different causes, but their symptoms are similar, so when accrue such the symptoms, it is necessary to visit the consultant doctor or review the (ED) in hospitals to receive appropriate treatment for the case that is diagnosed by the specialists, and if the infection continues without treatment, it may lead to infection chronic disease , the manual diagnosis to distinguish between acute bronchitis and acute asthma is cumbersome and takes a long time due to the large number of overlapping symptoms between the two diseases that may reach 12 common features to obtain accurate clinical examination by the specialist , many junior doctors in hospitals face great difficulty in making the right diagnosis and taking the appropriate decision therefore, specialists are consultation to in most cases , in recent years artificial intelligence techniques such as ML and DL have gained great and noticeable interest in the field of health care worldwide, where they have achieved rapid success, artificial intelligence techniques are considered one of the means that reduce pressure on all health institutions in many countries, especially those with limited income that lack medical equipment and devices[7],[8]. Huge amounts of medical data can be handled with a ML approach and speed up processing power and find patterns between various and multiple sources and increase predictive power ,they can also reveal the complexity and distinction between nonlinear sources of latent variance associated with disease and early disease detection especially diseases with overlapping symptoms[9] .

1.2 PROBLEM DEFINITION

The main problem is the overlap of symptoms between acute asthma and bronchitis that which are among the diseases of the Lower Respiratory Tract Infection (LRTI) that affects in preschool children (under 6 years), and the difficulty of differentiating between them by junior doctors and practitioners, due to the presence of 12 clinical features between the two diseases, most of which are overlapping, as junior doctors face many urgent cases, especially in (ED) and at different times that need to take the right decision to differentiate between two diseases to preserved the lives of children .

1.3 SIGNIFICANCE

The significant lies in the main and effective role of artificial intelligence techniques represented by ML and DL in the field of health care, especially in recent years through the development of health institutions around the world and building smart systems that meet the needs of citizens, especially in poor and low-income countries and provide the best services, among the health problems facing pre-school children are acute asthma and bronchitis because their weak physical strength and lower immunity compared to adults, many junior doctors find it difficult to diagnose diseases in children and make the right decision in order to receive appropriate treatment , the importance of the problem is to find ways and means of learning easy to implement and apply to distinguish between the two diseases through the introduction of ML algorithms and DL in the health aspect to find accurate and comprehensive solutions represented in early diagnosis of these diseases, as this will contribute to reducing time, effort and money especially in rural areas that lack medical devices and equipment, it will also promote the dissemination of electronic culture, motivate doctors and increase self-confidence to take appropriate decisions concerning the patient without referring to specialized doctors.

1.4 SCOPE

The scope of the study will focus to use of DL and ML methods by training algorithms that appropriate with our dataset to obtain a binary classification that differentiates between acute asthma and bronchitis for children under 6 years whose symptoms differ to adults. the study was conducted in Fallujah Teaching Hospital for Women and Children within the paediatric department, where the real data was collected during four months, starting in March 2022 to end of June 2022, and the percentage of data varied during the months according to the difference in children's immunity and weather fluctuations as high and low temperatures as well as other risk factors such as dust, smoke, fires, etc., where the consulting doctor had a great role in the process of collecting data and diagnosing cases and differentiate between them .

1.5 AIM AND OBJECTIVE

1.5.1 Aim

The aim of this study is to find a right and accurate diagnosis capable of differentiate between acute asthma and bronchitis in preschool children by using ML and DL and to avoid a wrong diagnosis that will sometimes lead to death due to the frequent overlap of symptoms between the two diseases , as the two diseases are sometimes confused by junior doctors and practitioners and this may cause some confusion in the diagnosis that may occur in emergency cases , in addition this study will be reduction in demand for medical equipment and devices in hospitals especially were capabilities are not available in addition to high accuracy and time reduction for junior doctors and practitioners , it will also contribute to enhancing the capabilities of the medical staff in hospitals and access to accurate diagnoses to distinguish between two diseases.

1.5.2 Objective

a. Obtaining an accurate diagnosis that differentiate between acute asthma and bronchitis that affects lower respiratory system in preschool children using several of ML and DL algorithms of the classification models to find the best results by collecting more data.
b. Helping the junior doctors by making the correct and accurate diagnosis before referring to the consultant paediatrician.
c. Accurate and rapid detection of early diseases in emergency cases.
d. Preserving lives of children.
e. Reduce time and effort and reduce the demand for devices and equipment inside the hospitals, especially in poor or low-income countries.

1.6 ORGANIZATION OF THESIS

The general structure of our study includes a general introduction to acute asthma and bronchitis that affects the lower respiratory tract infections in preschool children, where the introduction included an overview of the subject and what are the modern methods and techniques used for our topic by reviewing some of previous studies, which will be explained in detail in five other chapters.

Chapter 2 (Literature review): In this chapter, we will a comprehensive review of the previous literature that includes studies about acute asthma and bronchitis in preschool children in addition other diseases and how approaches ML and DL were used with an explanation of the methods of work and the use of algorithms and what results were obtained and discussed.
Chapter 3 (Concepts of ML and DL): In this chapter, the ML and DL concepts will be clarified, as well as their types and applications especially in the field of health care, and what successes they have achieved and the challenges, and the modern methods used in study.
Chapter 4 (Models proposed and methods): In this chapter, the software part will be explained, and the programming language used and many modern models for both ML and DL will be proposed that fit the collected dataset and how to process and clean it to be ready for training and testing in addition using the evaluation methods for each model and applying all methods to avoid errors.
Chapter 5 (Results and discussion, model selection): In this chapter, the results of each models used in ML and DL will be presented and discussed, in addition to comparing them and choosing the most appropriate model for our study.
Chapter 6 (Conclusions): It will be the last chapter of this study and will include a summary of the thesis with the results and the selection of the most efficient model as well as the future work that we propose.

2. LITERATURE REVIEW

A literature review is a process of survey scholarly references about previous studies of a particular topic, in which the reader can identify the gaps and relevant areas to current research by providing a comprehensive overview of current knowledge, In addition to making a critical analysis of those references by identifying the topics and discussions related to the topic of my thesis that were collected by previous authors, where the views on some important issues will be compared and cited, and the model studies will be highlighted about how our study of literature in general related , in addition, we will express our opinion by evaluating these data from a theoretical point of view .

2.1 MACHINE LEARNING APPROACH IN HEALTHCARE

Yao Tong et al. (2022) [10] , Performed a study to predict the future continuing of care (COC) by developing machine learning models for asthmatic patients and discovering associated factors, the dataset consisted of 31,724 cases to patients who received the health care from Washington University of Medicine for nine years , started in January 2011 to December 2018, where the building of the machine learning model relied on examining 128 features with 10-fold cross-validations of the model , Several models of ML have been used Baseline KNN , Naive Bayes , (SVM) , Random forest , XGBoost (extreme Gradient Boosting) , where the highest accuracy was 88.20 % for the XGBoost model, however, they used the (ROC), which was 0.96% , and the average f1 score of 0.86 %, the model can facilitate future clinical decisions and improve hospital management outcomes . Although the accuracy is fine, researchers should have trained other types of algorithms to obtain higher accuracy. Katy Stokes and his colleagues (2021) [2], used machine learning methods to diagnose between pneumonia and bronchitis for a group of middle-income patients in Sarajevo Hospital and through which data were collected for the infected patients, which consisted of 4,500 real cases that were clinically examined by specialized doctors, dataset consisted of 3000 cases of pneumonia and 1500 cases Bronchitis started in October 2017 to December 2018 , the dataset were divided into 60 % for training and 40% for testing and during the study, three algorithms for ML were tested, namely, logistic regression, decision tree and support vector machine (SVM) , after being developed by researchers and comparing them , the decision tree got a highest accuracy of 93% with AUC . Although the number of the data set was good, it was not close in terms of percentage in addition, the researchers did not use other models to find a higher accuracy. Junfeng Peng and his colleagues (2020) [11], developed a new C5.0 decision tree classifier for predicting the diagnosis of acute exacerbation chronic obstructive pulmonary disease (AECOPD) . a real dataset of 410 cases collected at Sun Yat Sen University Third Hospital, which included 28 features, including medical and vital signs , the cases were clinically examined by Specialists: these cases varied to 208 cases with COPD, while 202 cases with acute exacerbation of COPD , the data set was divided into 80% training and 20% testing, where the accuracy of the model C5.0 was 80.3%, with confidence interval 95% which exceeded the rest of the models used such as classification and regression tree (CART) , ID3 (Iterative Dichotomier 3) , C4.5. Despite the small number of data, the accuracy was not high. Yoshihiko Raita et al. (2020) [12], proposed a study aimed at developing ML models by predicting the severity of bronchiolitis for prospective cases of infants under one year , the real data of 1016 prospective cases were collected for a period of three years, starting from 2011 to 2014 distributed over three seasons (from November to April), where they were clinically examined by specialized doctors according to the guidelines of the American Academy of Paediatrics, where most of the cases were shortness of breath (SOB) , the average of age patients was 3.2 months, and the percentage of females was 42% and the length of stay in the hospital for the disease cases was 0 to 60 days , searchers developed four machine learning models: Lasso regularization with Logistic Regression ,elastic net regularization with Logistic Regression , Random Forest , Gradient boosted decision tree , after comparing the four models, Gradient boosted decision tree outperformed with an accuracy of 95% and (AUC) of 88% . The researchers' idea was good by developing models and obtaining high accuracy after collecting real prospective data, but the percent of (AUC) does not correspond to some degree of accuracy. Rahat Ullah et al.(2019) [13], presented a study on predicting the risk of developing asthma by evaluating blood samples of patients using machine learning models, where the data set consisted of 52 healthy blood control cases and 150 Raman spectra cases, the data were obtained from the Hospital of Medical Sciences Islamabad in Pakistan institute , three results of machine learning algorithm were compared , Artificial Neural Network , Random Forest , SVM , Algorithm (SVM) got the highest accuracy of 94.1% with an (ROC) . The data set was small; However, the accuracy was excellent.

Nida Shahid and her colleagues (2019) [14] , conducted a basic review on organizational decision-making in the field of health care using artificial neural network applications. the review included a survey of 3,397 articles in six databases related to computer science and health administration , coverage that were published for the period from 1997 to 2018, in 24 countries the objectives, methodology, context and analysis were clarified, there were 80 articles according to the inclusion criteria , published (26 articles) to authors from the United States , Feed-forward Networks (FFN) , (25 articles) types of models ANN used included ANN (36 articles), or hybrid models (23 articles) , reported accuracy percentage varied from 50% to 100% , informed decision-making of the plurality of ANN at the micro level (61 articles), between health care providers and patients . were deployed for intra-organizational of fewer ANN (mid - level, 29 articles), and policy or inter-organizational system, (high- level, 10 articles), the study determined the mechanism of artificial neural networks by reviewing the main characteristics and drivers and adopting this technology through making organizational decisions for health care. The review conducted by the researchers was good because it included many articles published in international accounts. Dimitris Spathis et al. (2019) [15] , discussed the support of clinical decision systems in the field of health care through prevention and diagnosis of respiratory diseases such as chronic obstructive pulmonary disease (COPD) and asthma, where the study was conducted in Greece for patients who visited a clinic of a consultant doctor in the outskirts of Thessaloniki, the number of cases was 132 cases for the period from 2014-2015 and was each patient has 22 different values that included many types of symptoms , several machine learning models have been implemented , Logistic Regression , SVM , K-NN , Random Forest , Artificial Neural Network , Naive Bayes , Decision Tree , classifier of Random Forest outperformed other techniques with an accuracy 97.7 % . Despite the small number of data, the accuracy was high, where the researchers used many algorithms to compare and show the highest results. Julie L and his colleagues (2020) [16] , presented a study that included the use of ML models to predict low-severity asthma and the use of limited and small samples, where models were developed to predict the analysis of the data set of the health of the child with asthma, the data set consisted of 50212 cases that were obtained from websites on the prevention of diseases of children With 23 variants per affected child such as Sex, (SOB) , Allergies . the results of four models of ML were compared, Decision Tree, Random Forest, Naive Bayes, Linear Regression, K-NN, classifier of random forest resulted in highest prediction an accuracy of (90.9%). Alberto et al. (2018)[17], presented a study based on the comparison of distances for the K - NN method in diagnosing heart diseases, where the study was carried out by two methods, the first is calculating five distances for KNN and the second is calculating three distances , the final results showed that calculating the distance of three nearest neighbours using Euclid, Manhattan and Mahalanobis was 85% for the data set collected from UCI heart Cleveland data . Yun-Chung Liu and his colleagues (2022) [18] , conducted a study to develop ML algorithms to predict patient g intensive care unit (ICU) needs for children with pneumonia within a period of 24 hours and make appropriate decisions by identifying key indicators of patients and evaluating their performance , the number of cases reached 8,464 , searchers compared clinically pertinent features between cases without ICU care with early ICU transfer cases , the researchers used three models of machine learning XGB, Logistic Regression and Random Forest ,algorithm of Random Forest done the best performance(accuracy of 0.936% sensitivity 0.94 AUC 0.99 %) .Tingting Cao et al. (2022)[19], conducted a study to predict Henoch Schönlein of purpura nephritis (HSPN). in children, where real data were collected for 240 infected cases who were clinically examined inside the Children's Hospital for the period from October 2019 to December 2021 with 10 important features such as age, abdominal pain, and measurement of the amount of protein in the urine within 24 hours. Males, 126 cases and 114 females, where the complex cases of 153 cases were difficult to diagnose, algorithm of XGBoost prediction model was built and the results was (AUC) of the training dataset (0.895, 95% CI: 0.827-0.963) and testing dataset models were similar (0.870, 95% CI: 0.799-0.941). Yinhe Feng et al (2021)[20] , presented a study on a survey of artificial intelligence and ML and its four applications in asthma and COPD through classification, diagnosis, assessment, and monitoring , where researchers demonstrate the possibility of integrating large and heterogeneous medical data and thus facilitating decision-making and guidance by clinicians regarding clinical practices, these techniques can also be used to provide treatment guidance and analyse different responses to treatment . In the end, the researchers emphasized that machine learning and artificial intelligence techniques cannot replace doctors in the diagnosis and treatment of asthma and COPD, as they need large samples with external data sources for examination, verification, and comparison with the results. Sushruta Mishra et al. (2020) [21], presented a study on introducing a hybrid algorithm of type Enhanced Adaptive Genetic Algorithm (EAGA) with (MLP) for diagnosing diabetes in a group of patients ,the data was obtained from websites where the dataset divided into 60% training and 40% testing , (EAGA-MLP) model which proposed outperformed of accuracy of classification of 97.76% with delay of only 1.12 of sect he least latency to perform the classification process. Sohely Jahan et al. (2021)[22], suggested several ways for machine learning models to predict cervical cancer in women using risk factors. Researchers used 8 types of classification algorithms for early diagnosis and detection of the disease , Support Vector Classifier (SVC) ,Perceptron (MLP), Random Forest , K-NN, AdaBoost , Decision Tree , SVM , Gradient Boosting, Logistic Regression , data were collected in Venezuela, of 2017 from the University of California in Hospital de Caracas , the number of cases was 858 real cases that were clinically examined by specialists, in addition to 32 features for each case , MLP algorithm outperformed with accuracy 98.10 % . Ali Al Bataineh and his colleagues (2022) [23], developed a set of smart systems by using ML algorithms such as Decision Tree Classifier , XGB Classifier ,Extra Trees Classifier , KNN Classifier , GaussianNB Classifier , Gradient Boosting Classifier , Logistic Regression Classifier , MLP Classifier BP , Random Forest Classifier , SVM Classifier to predict the diagnosis of cardiovascular diseases through Cleveland Heart Disease dataset that are publicly available on websites. searchers proposed MLP neural network trained with a particle swarm optimization (PSO) using 13 features by which to distinguish the presence or absence of disease for each case, the data set was divided into 70% training and 30% testing, where the model of MLP exceeded the rest of the used ML models with an accuracy of 84.61. Gang Yu and his colleagues (2020) [24] discussed improving artificial intelligence models to identify asthma quickly and accurately in paediatric general departments and reduce antibiotic use. they were able to obtain the real dataset of asthmatic patients from a children hospital affiliated with Zhejiang Medical College in China, examined by specialists. Clinically, the data set is consisted of by two test parts: the first part consisted of 325 positive asthma cases with 428 negative asthma cases related to lung diseases, while the second part consisted of 2,123 cases from sections not related to lung diseases, including 1786 negative cases with 373 positive cases , Four ML models were used , Naïve Bayes , SVM ,Logistic Regression ,CatBoost , the CatBoost models model outperformed over models on both test by an accuracy of 84.7% with (AUC) of 90.9% on test 1, and an accuracy of 96.7% with an (AUC) of 98.1% on test 2 . OLUFUNKE C. OLAYEMI and his colleagues (2019) [6] , conducted a study on the use of ML techniques to diagnose the presence of lower respiratory tract infection (LRTI) in infants, where the real data set that was examined by specialists was obtained from the southwestern Hospital in Ondo State was pre-processed and obtain the relevant attributes , which is in Nigeria, where the number of 702 prospective cases , the researchers used two models of machine learning, K-NN and Naïve Bayes with all features (18) accrued 94.25% and 94.43%, respectively , Naïve Bayes with dataset based feature selection process shows accuracy of 99.60% while K- NN shows 94.35% with 10 features , the Comparative results show that Naïve Bayes selection method works Stronger and best than others with the dataset based feature.

2.2 DEEP LEARNING APPROACH IN HEALTHCARE

Lamis F. Samhan and his colleagues (2022) [25] , conducted a study reviewing the methods and techniques used to classify Alzheimer's disease using DL models, specifically convolutional neural networks (CNN) , the researchers used a data set obtained from the Kaggle website, which consisted of 10,432 brain images of affected people to be categorized by the presence or absence of the disease , the data divided into training data at 70% and testing data at 30%. After training, the researchers had a model accuracy of 97% with 0.0832 Validating loss, 0.0012 training loss. Jahan and his colleagues (2021) [26] , conducted a study on the application of (CNN) algorithm to diagnose skin cancer, whereby early detection of the disease and isolation of patients to prevent the spread of infection , the searchers use of ML models such as decision trees , random forests , GBT and was compared with the DL algorithm, where the final results showed the superiority of (CNN)with an accuracy of 88% . Although the researchers used several algorithms of ML and DL, they did not use techniques to avoid overfitting during training and testing. Arpan Srivastava and his colleagues (2021) [27], presented a study aimed at applying DL models through the use of (CNN) model that help provide accurate and detailed information in the analysis of audio and medical data through sounds that help medical personnel in detecting COPD, where the model can be classified into three categories Severe, medium and mild , the data set was obtained during the (ICBHI, 2017) , after ML models such as KNN and SVM are compared with DL model the final result was applying (CNN) algorithm to the data, amount of accuracy was 93% with ten split and applied K-fold Cross-Validation to obtain a high performan .Charles Bales and colleagues (2020) [28], presented a low-complexity study using (CNN) to examine respiratory infections in children by detecting intraoral coughing, converting sound into Mel-Spectrograms, and diagnosing three potential diseases: bronchiolitis, pertussis, and bronchitis , the researchers obtained a data set from open sources on the net, the number of coughing cases was 933, while there were 933 cases of non-coughing, (CNN) achieved an accuracy of 89% . Chen Chen and his colleagues (2019) [29], proposed a method that combines two models of DL, (CNN) with (LSTM) , to detect heart disease by automatically identifying six ECG signals , searchers used dataset from PhysioNet website in addition three sets of ECG data were used for comparison and evaluation , the final results showed an accuracy 99.32 . Serkan Kiranyaz et al.(2021) [30] , presented a comprehensive study on (1D CNN) and their applications, and the researchers refer that the first to propose this network was Kiranyaz and his colleagues in 2015 and applied it to the ECG signals of patients, and it was small and has the ability to adapt, in addition to that the real time in it is low cost on the contrary exactly with (2D CNN) , where it has a large bifurcation and operational time complexity, and 1D CNN can easily train and achieve high performance by providing a minimum computational complexity and solve problems related to classification, such as classification of paediatric cardiorespiratory diseases, etc., having the ability to merge the features of extension and application and make them into a single adaptive environment . Parvathaneni Naga Srinivasu and his colleagues (2021) [31], proposed a computerized study to classify skin diseases by merge deep learning models using (LSTM) with MobileNet V2 , where the LSTM works to retain feature information from the previous state and improve the performance of the model ,the researchers obtained the dataset on April 17, 2021, from the Kaggle website where the data set consisted of 10,000 images of different types of skin diseases for people around the world, the final results showed effectiveness in classifying skin diseases by merging the two models, and the accuracy reached 85%, which enhanced the predictive power of classification, in addition, the two combined models were compared with the other models Such as (CNN) and Fine-Tuned Neural Networks (FTNN) , and it was noted that the first model is the best in terms of classification, analysis, and detection of tumours. Pham, Tuan D. presented a research paper(2021) [32] , on a strong scientific classification based on time and physiological series through the use of a developed model of TF - TS LSTM for place, time and frequency, because the traditional solutions of LSTM leads to a low level of accuracy, where the study was conducted on patients with Parkinson's disease through using walking data, the idea was to create identical temporal features of place, time and frequency instead of watching the time, the data set was obtained from the website PhysioNet , TF - TS LSTM surpassed the traditional LSTM 94%. Fatih Demir et al. (2021) [33], presented a powerful approach based on deep learning models by integrating LSTM networks with Recurrent

Convolutional Neural Network (R-NN) , this approach helps to detect different eye diseases through fundus images and deep feature extraction. the dataset was obtained through ODIR data containing coloured fundus images of the eyes of patients volunteers were collected for the right and left side of the eyes, which were formed by special cameras and saved in different special sizes and in JBG format, the number of samples was 6426 samples from eight categories and the NCAR algorithm was applied to extract the features from them , the researchers applied the proposed approach with NCAR algorithm and SVM classifier , the final results showed an accuracy of 89.54% with an AUC of 97%. Balaji E and his colleagues (2021) [34], presented a new structure of deep learning models for early and non-invasive diagnosis through the use of LSTM networks technologies to classify the intensity of Parkinson's disease through the walking cycle where the algorithm learn long -term temporal dependencies On the contrary machine learning models , because it solves the vanishing gradient problem , researchers obtained the collection of data from the Physionet , which contains VGRF data, dataset amounts to 13,000 samples for two minutes of walking, the researchers made two ways in the classification, the first classification of distinguishing between Parkinson's disease and health control and the second method is the multiple classification to distinguish between the severity of type Parkinson's disease , the final results showed that the binary classification, with an accuracy of 98.6%, was superior to the multiple classification, with an accuracy of 96.6% , the researchers also applied the ADEM optimizer in both models with a 3.4% accuracy improvement , comparison between the techniques of related. Mohamed G. El-Shafieyand his colleagues (2021) [35] , proposed to build a hybrid model of deep learning models through using Bidirectional LSTM with 1D CNN to predict heart disease through the binary classification for the presence or absence of disease ,the researchers used the data set from the UCI ML library, which contains two parts : the first section of Statlog contains 120 cardiac records while Cleveland contains150 records , the dataset is divided into 70% training and 30% testing , final results of the proposed approach achieved in the binary classification was an accuracy of 89.01% and 82.72% respectively to the datasets of Cleveland and Statlog , in addition used AUC and ROC . Mohamed Djerioui and his colleagues (2020) [36] , proposed an intelligent system based on LSTM techniques to predict, classify, and compare heart disease with multi-layers perceptron (MLP) , the researchers used the cardiovascular data set at the University of California USA, which was collected from the data warehouse for machine learning , the database consisted of 303 patient records with 76 features for each patient, but the researchers used only 14 of them to predict the presence or absence of disease, out of a total of 303 records were there are 165 actual patients while 138 have no disease. the model proposed by the researchers achieved an accuracy of 99.8% over the MLP which achieved an accuracy of 94.73 %. By reviewing the literature for research, we noticed that most researchers are moving towards health care, especially in recent years, through the application of artificial intelligence techniques within that field, which a great achievement that has great benefit to societies, especially the poor or limited income , through early detection of diseases that It is difficult to diagnose and receive treatment before the disease worsens, especially in children , some researchers used machine learning techniques by applying algorithms related to the nature of the data, and the most used were (KNN, Decision Tree , Random Forest , MLP , Naive Byes .. etc.), which achieved good results despite the presence of some gaps such as small of data or the use of false data for patients from websites and some of them use only two models or not use the model that fits with the size and nature of the data, whether it is text or pictorial data, and all of this may negatively affect the performance model in terms of classification and diagnosis, as for DL models, the review of sources was better than ML models due to the presence of newer techniques that allow algorithms to deal with, analyse and filter data and find appropriate solutions for them, especially problems related to complex diseases through classification and early detection, some of them contained a binary classification and the other contained a multiple classification according to the severity and type of disease, such as respiratory diseases in children, heart diseases, skin diseases, and others the most widely used are (CNNs) and(LSTM) , due to the presence of many features in these algorithms, such as data pre-processing, especially as they deal with large and complex data and give the best solutions, making them easy to implement and apply. there are some gaps by the researchers while applying algorithms to the diseases such as not avoiding overfitting during training and testing, not using dropout techniques for layers or batch normalization or the error rate in confusion matrices, however, the performance rate of most of the algorithms was good and achieved successes in the field of health care.

3. CONCEPTS OF ML AND DL

3.1 ML AND DL (CONCEPTS, TYPES, APPLICATIONS)

ML is one of the branches artificial intelligence which make the machines to learn through experiments and contains many algorithms that can extract patterns from the data and link those patterns to categories of samples within the data and do predictions with least human intervention the term machine learning differs to artificial intelligence, where machines that simulate intelligence are made by artificial intelligence techniques, while algorithms are built and configured in a programmed manner through ML , DL is a branch of ML, which is an artificial neural network consisting of three or more layers that simulates the behaviour of the human brain , it allows learning from large data sets , DL algorithms learn and process large, unstructured data sets like texts, images, and distinguish between things, unlike ML [37].The Figure 3.1 explain the tasks of ML and DL within in artificial intelligence .

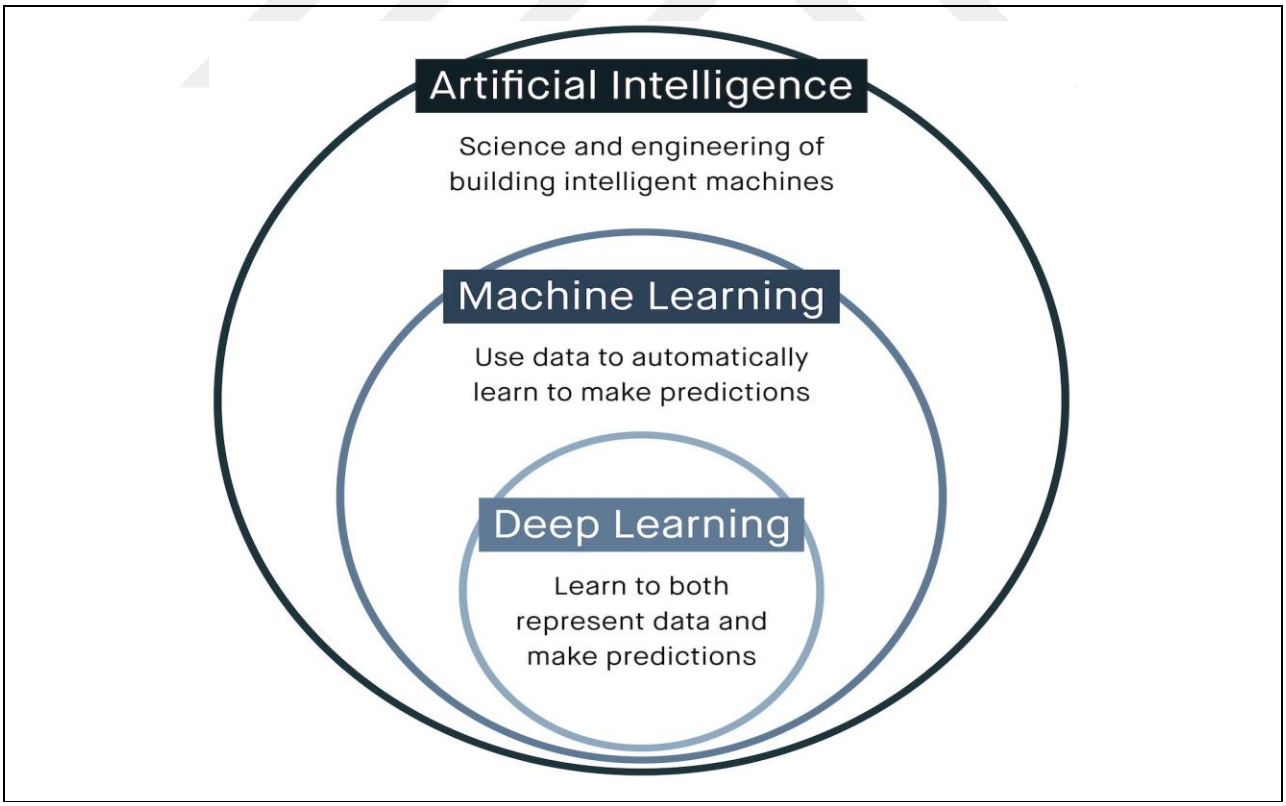

Figure 3.1: Tasks of ML and DL

As these algorithms use of sensitive methods to learn from the data without relying on prior procedures in addition, it improves its performance while increasing the number of data adaptively, ML approaches can be identified based on the problems that the algorithms face, such as their interaction with the input data types and the environment, in general ML works by forming algorithms to create a specific model after training the data set and creating a developed model to make predictions[38]. As in Figure 3.2.

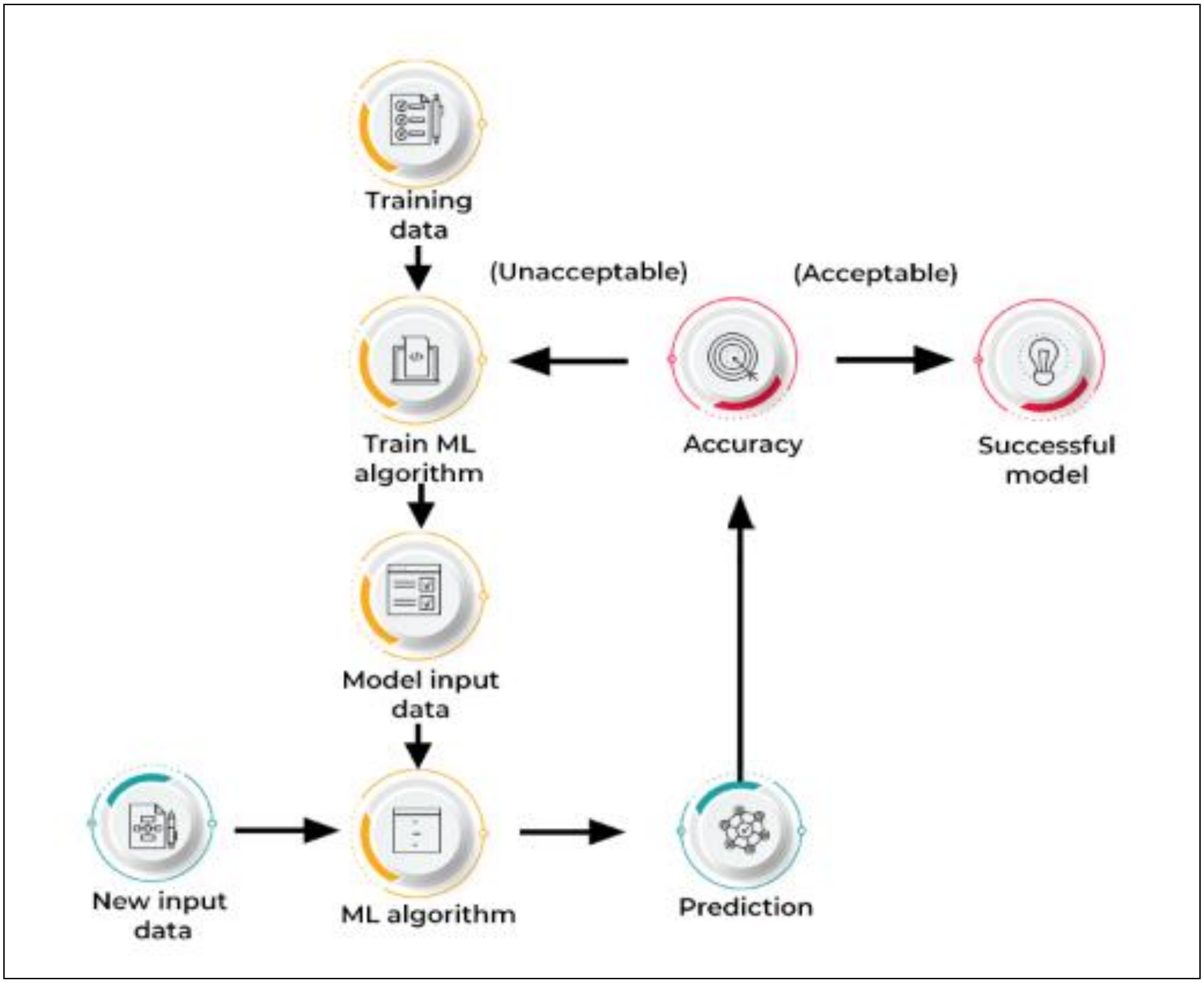

Figure 3.2: Method work of ML

The above Figure (3.2) shows a general, high-level model of ML models that includes many variables, factors, and steps, in addition, the prediction can be verified through the accuracy of model (acceptable, unacceptable), and the required accuracy can be achieved using a training data set that is enhanced frequent.

3.2 ML TYPES

ML can be categorized into four basic parts, supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning, as in Figure 3.3.

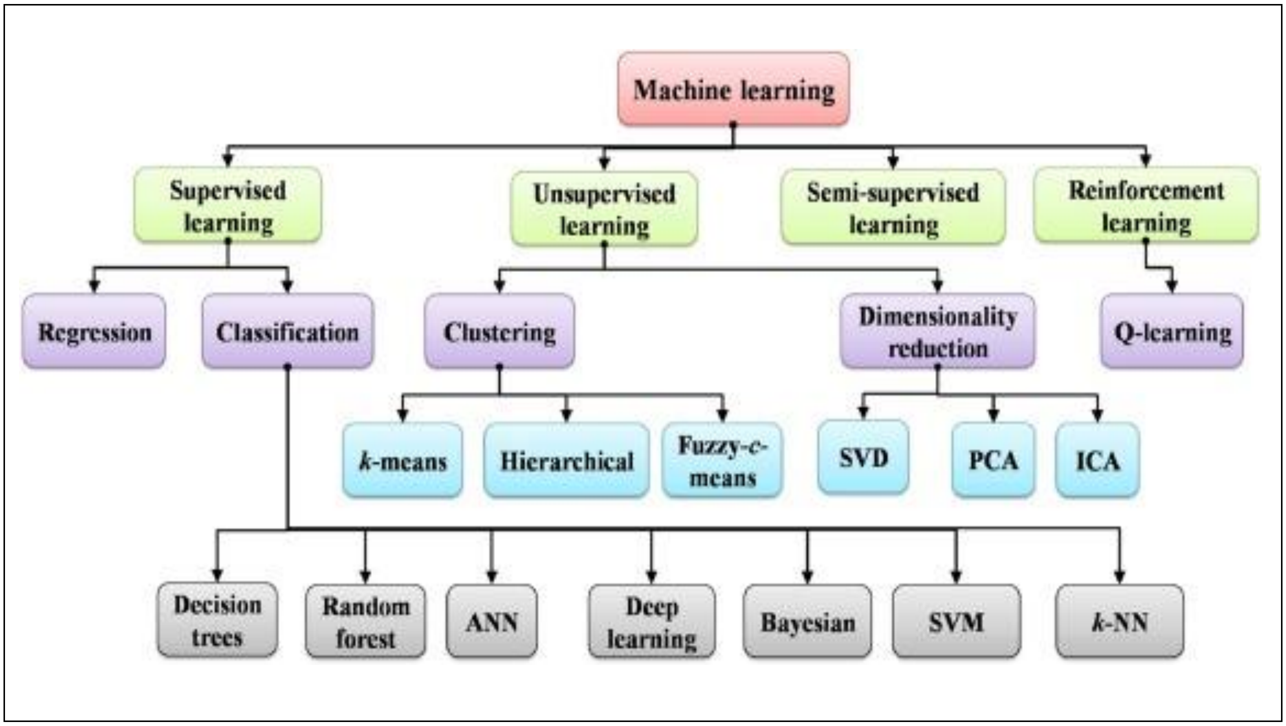

Figure 3.3: ML types

3.2.1 Supervised Learning

In this type, the machines are trained on the dataset in terms of inputs and outputs, naming the results, and enabling the machines to make predictions based on the pre-training of the data (supervised) through testing ,the main objective of ML lies in identifying the input variable (A) with the output variable (B) to Explicitly and repeatedly formulating the model, in addition to pre-processing the data, which is considered the most important step and may affect to the model performance , as the data is unified, cleaned, limited to the required variables, and other variables are removed and evaluate the model through validations using receiver operating curve (ROC) , confusion matrices to obtaining final model with the best parameters and a high degree of predictability, predicting such as the presence or absence of disease in children based on the Inputs and outputs that have been trained so the dataset must be clear and accurate [39]. There are two types of supervised learning as Figure 3.4.

a. Classification: In this class, algorithms address classification problems, and the variable output is categorical, such as presence/absence of disease if the classification is binary while severe, moderate, or mild if the classification is multiple.
b. Regression: Algorithms in this class work on regression problems address by forming a linear relationship for input and output continuous variables such as predict analysing market trends and weather condition.

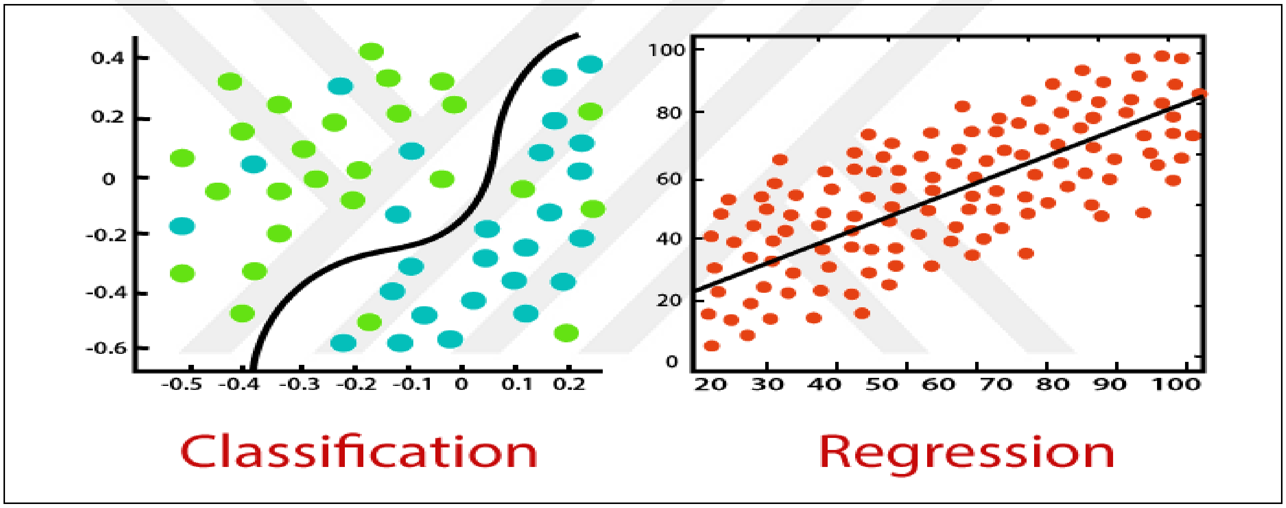

Figure 3.4: Classification Vs Regression

3.2.1.1 Classification algorithms

K - nearest neighbours :

It is an algorithm that is easy to interpret, implement, and is versatile because it does not require setting many parameters. Therefore, it is considered one of the most accurate and time-efficient algorithms among other ML models , it is used in both classification and regression, as it assumes the existence of similar data close to each other [40]. As in Figure 3.5 ,this assumption is true when the distance or proximity between similar data is minimal, as the distance is calculated by several methods, including the Euclidean distance as show in Figure 3.6 , which is commonly used and the equation 3.1 show the distance between A and B, in addition to the Manhattan, and Minkowski distance [41] .

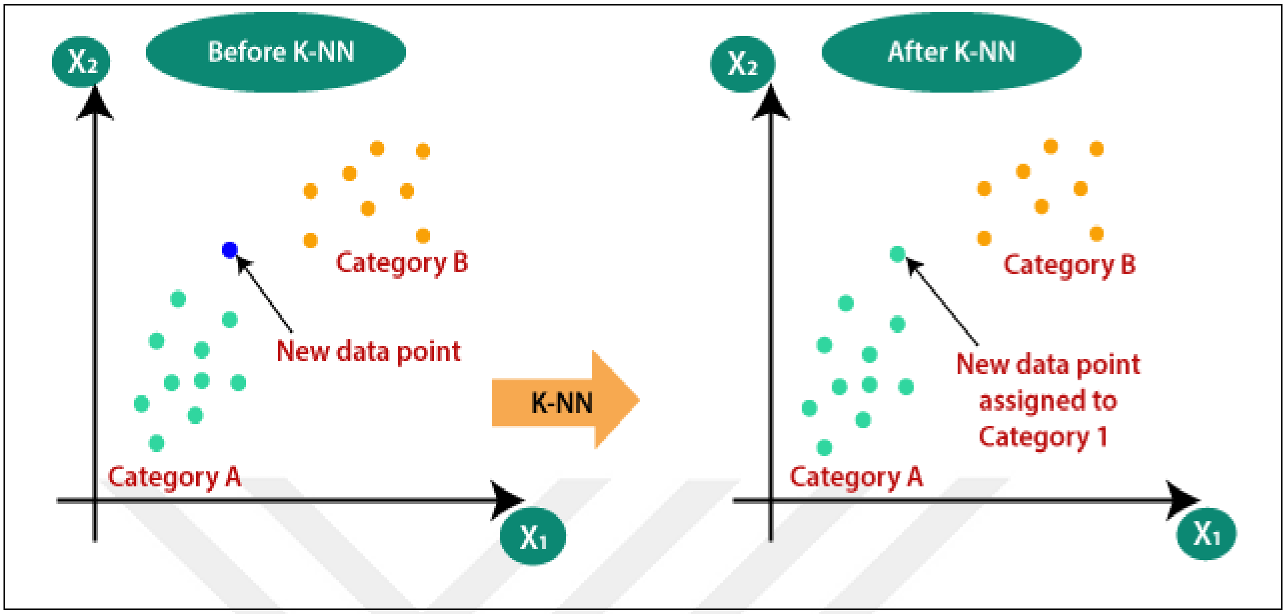

Figure 3.5: K-NN identify the category

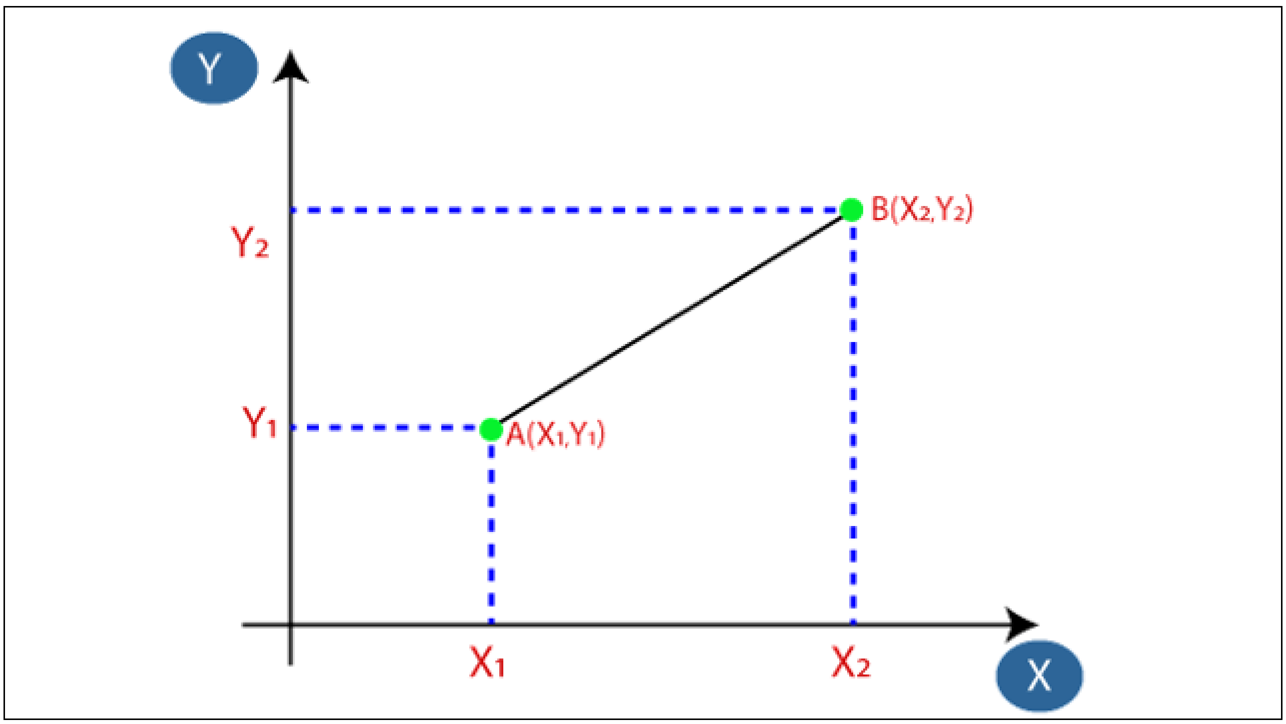

Figure 3.6: Euclidean distance

$$distance\ between\ \ A\ and\ B\ \ = \sqrt{(X2 - X1)2 + (Y2 - Y1)2} \qquad (3.1)$$

There are several disadvantages of K-NN algorithm [42].

a. It calculates the distance between points for all training data, so the cost will be high.
b. It is complicated at some time because it depends on setting K value always.
c. As the number of examples and predictions increases, the algorithm becomes noticeably slow.
d. When you deal with high-dimensional data, it will lead to a decrease in accuracy, so it can be used to classify multiple diseases that show similar symptoms, as well as in image and video recognition and handwriting detection ,credit classification by using feature similarity.

Decision tree :

A decision tree is used to predict a categorical variable , the decision space is divided into smaller and smaller regions and then named (recursive partitioning) where it can adapt to the clinical context easily and interpret the best , It starts from the root of the tree to predict the category name for a record and compares the value of the root attribute with the attribute of the record it follows the section corresponding to that value and then moves to the next nodes , decision trees classify data from the bottom of the tree (the root) into some leaf / terminal with the leaf/terminal node to providing the classification , every node in the tree do as a test status for some attribute, and each verge descended from the node match to the potential answers for the test status as Figure 3.7 , decision tree can be used to solve regression problems as well [43].

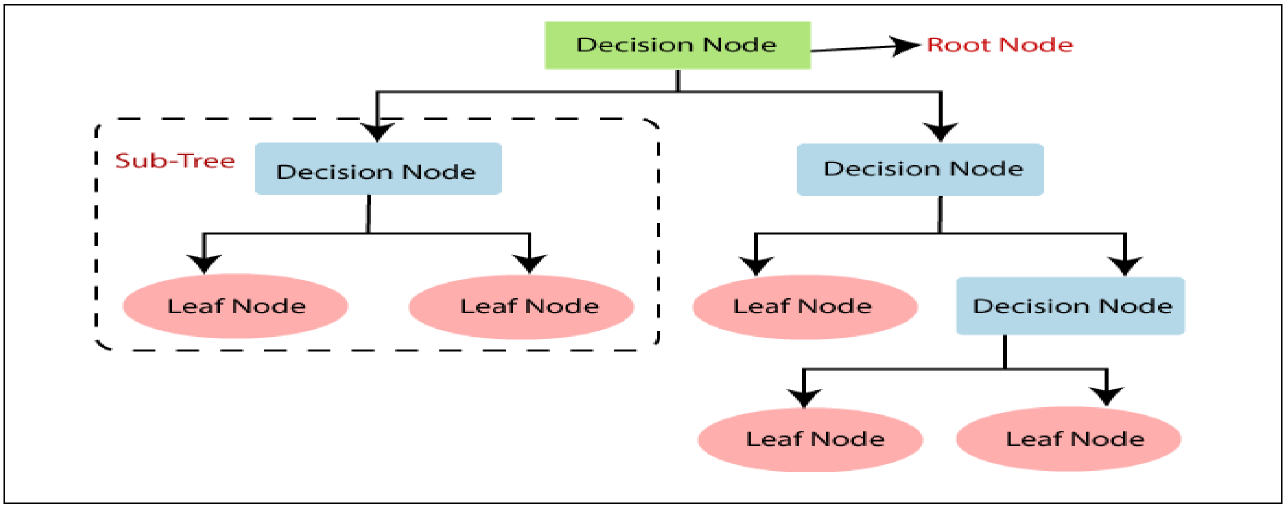

Figure 3.7: Decision tree work

The decision of strategic divisions affects the accuracy of the tree significantly, as the decision criteria differ for regression and classification trees, decision trees use several algorithms to divide the nodes into two or more sub nodes, and the more sub-nodes are created, increased the homogeneity of the resulting sub-nodes and the choice of algorithm depends on the type of target changes, there are several algorithms used in the decision tree:

a. Iterative Dichotomiser 3 (ID3).
b. successor of ID3 (C4.5).
c. Classification And Regression Tree (CART).
d. Chi-square automatic interaction detection (CHAID).
e. multivariate adaptive regression splines (MARS).

The degree of system chaos in the decision tree is reduced by selecting and dividing the nodes to obtain an optimal classification, where the degree of chaos is quantified as in the two equations 3.2 and 3.3

$$Gini\ (n) = 1 - \sum P_i^2 \qquad (3.2)$$

$$Entropy\ (n) = -\sum p_i \ {}_{\mathrm{Log}\,2\,(p_i)} \qquad (3.3)$$

(n) represents a sample of data, and (p_i) represents the frequency of class i in sample n

Both Gini and Entropy ensure a high-accuracy tree with small variance. entropy uses a complex logarithmic process to measure the chaos of the system, while Gini is used to evaluate splits in the data set (cost function) computed by subtracting one from the sum of the squared probabilities of each category as in equation , [44].There are many advantages and disadvantages in a decision tree as in table 3.1 [45] .

Table 3.1: Advantages and Disadvantages decision tree

Advantages of DT	Disadvantages of DT
Easy to interpret and use	High of complexity
It does not require data normalization	The nature unstable
Specific values are assigned to each problem	Consumes more memory
In the event of missing data, it does not significantly affect the process of building the model	Unsuitable for applying predicting and regression continuous values.

Artificial neural networks :

They are artificial networks that are inspired by the biological neural networks of the human brain and simulate it through mathematical operations , the human brain contains many neurons interconnected with each other as in Figure 3.8 , artificial neural networks also contain many cells interconnected with each other within different layers called nodes as Figure 3.9 [46].

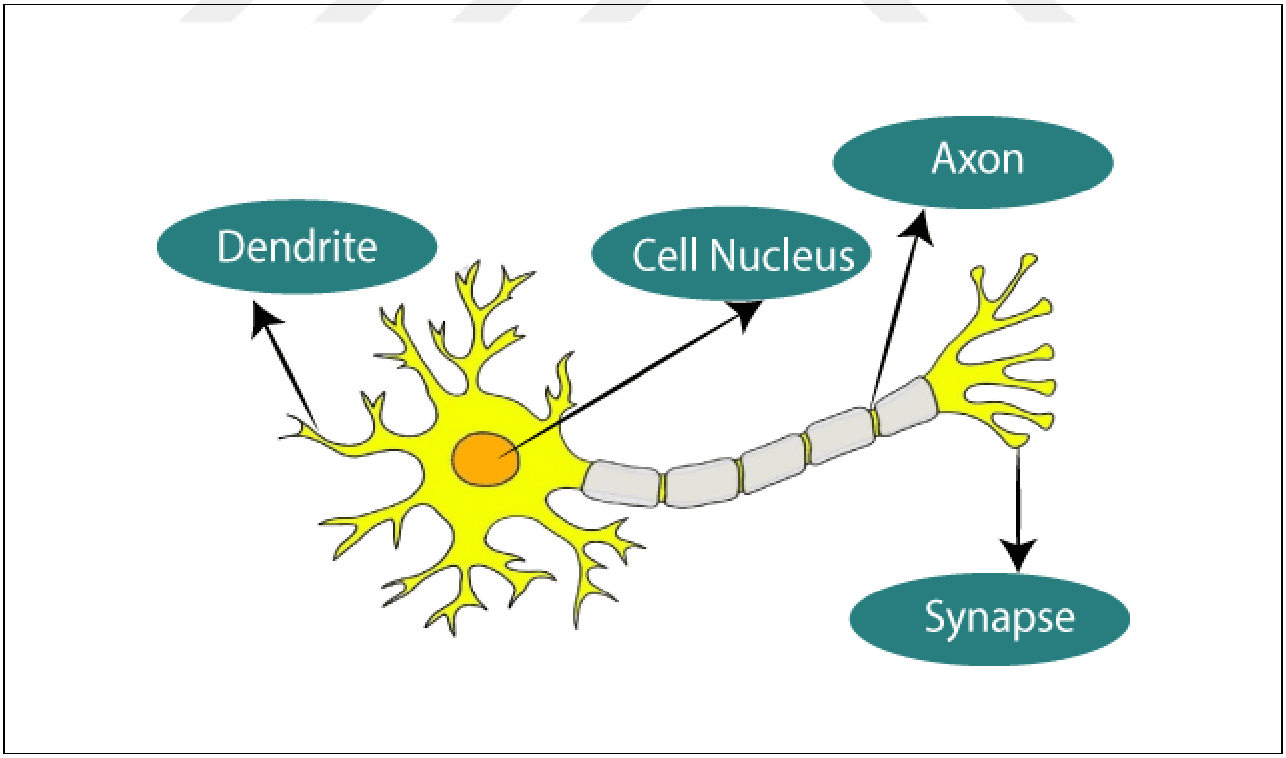

Figure 3.8: Biological neural network

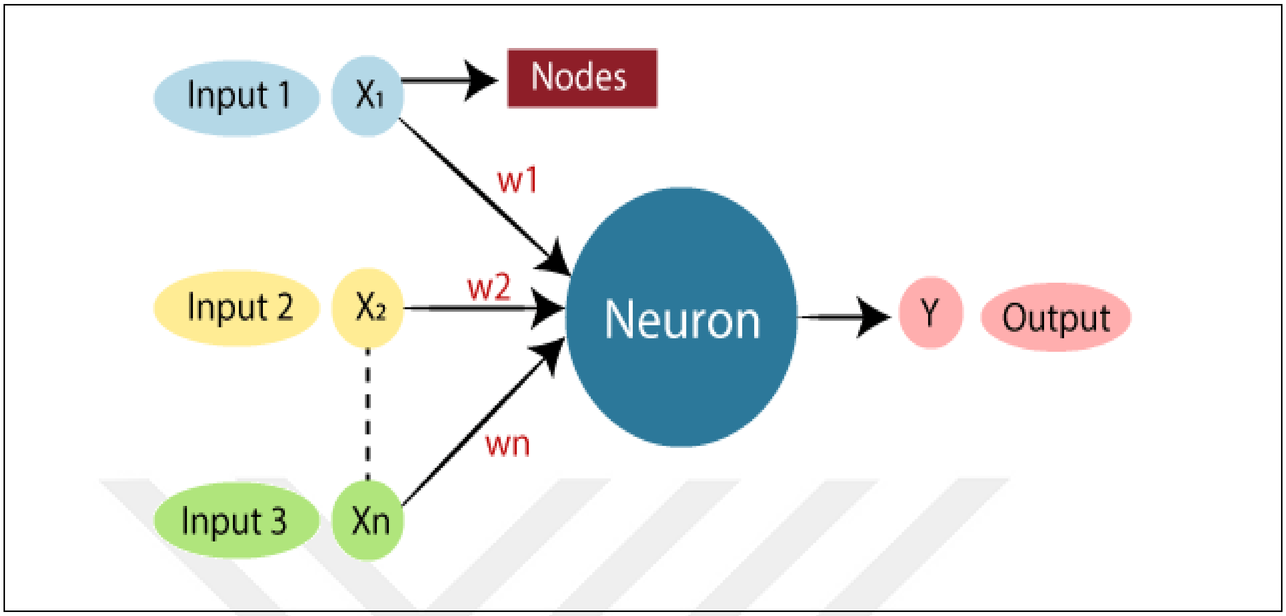

Figure 3.9: Artificial neural network

Thus, a relationship can be found between the artificial neural network and the biological neural network, as shown in the table3.2 .

Table 3.2: Relationship between artificial neural network and biological neural network

Artificial Neural Network	Biological Neural Network
Inputs	Dendrites
Cell nucleus	Nodes
Weights	Synapse
Output	Axon

Artificial neural networks simulate the human brain by making computers understand things, make decisions and behave in a manner similar to how interconnected brain cells think Where there are 1,000 billion neurons inside the human brain and each cell has a link point of 1,000 and 100,000 and the information is stored and distributed in a balanced memory and recalled when necessary, so that the human brain can process that data in a very amazing way , OR gate can be a simple example to understand the working of a neural network based on variable inputs which it takes two inputs (0 and 1), so if one or both of the two input are (ON) , the output will be (ON) , while in the case that the two inputs are (OFF), then the output will (OFF) , In recent years, Neural networks have become popular for use in the fields of health care, information security, and big data processing, because they have great advantages in the speed of processing large data, efficiency and accuracy in performance, and they have the ability to deal with complex problems in medical and agricultural sciences, security, education, management and others an artificial neural network consists of three main layers as in Figure 3.10 , [47] .

a. Input layer: It is the first layer that accepts data in many formats and passes it to other network.
b. Hidden layer: It is one or more hidden layers located between the input and output layer whose function is to extract hidden features and patterns through arithmetic operations and is responsible for network complexity and performance.
c. Output layer: It is the last layer of the network that contains the result after being passed and processed in the hidden layer and calculating the weighted sum of the inputs.

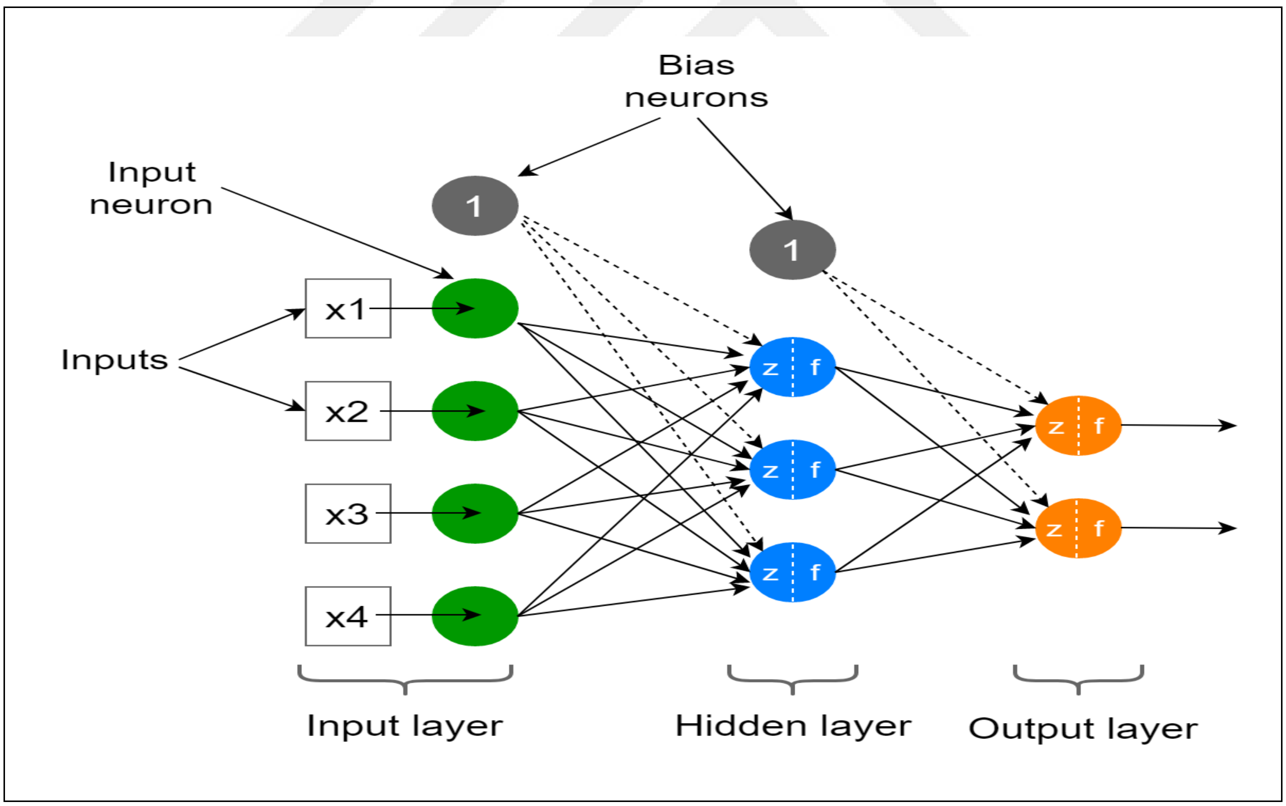

Figure 3.10: ANN layers

The artificial neural network works by receiving the input signal from the external interface in the form of an image and a pattern where the inputs are set by the symbols x (n) mathematically as in

Figure 3.11 and each n has a number of inputs as equation 3.4, then each input is multiplied by the weights that correspond to it , weights are an important part within the neural network because they solve specific problems and have the strength of the interconnection between neurons. after that, the computing unit stores all the weighted inputs inside it , when the sum of the weights is zero then the output will be nonzero or something else to extend the range by adding bias to the inputs , while if the weight is equal to 1, this means that the total weighted inputs will be within zero to positive infinity to maintain the desired specified value, in addition, a certain maximum value is measured and through the activation function, passed the total weighted input ,the activation function is responsible for achieving the desired output through the set of transfer functions used , there are different types of activation functions, including linear and non-linear, and the most used are tan hyperbolic , sigmoid, Relu , as in Figure 3.12 , [48][23].

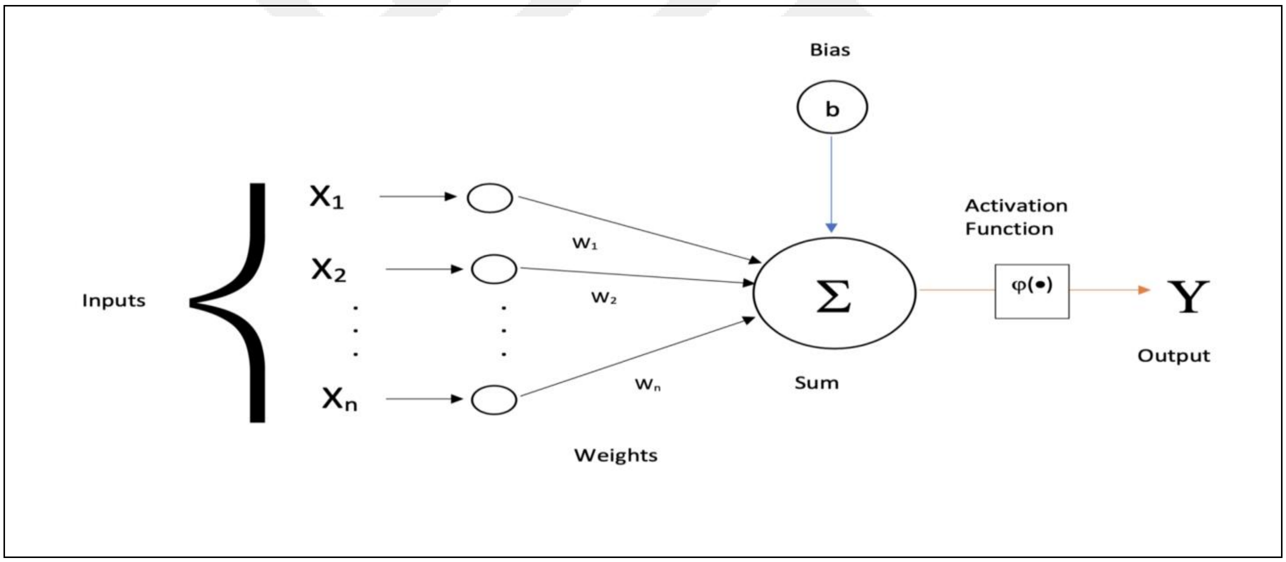

Figure 3.11: ANN Work

$$y = f\left(\sum_{i=1}^{n} wi * xi + b\right) \qquad (3.4)$$

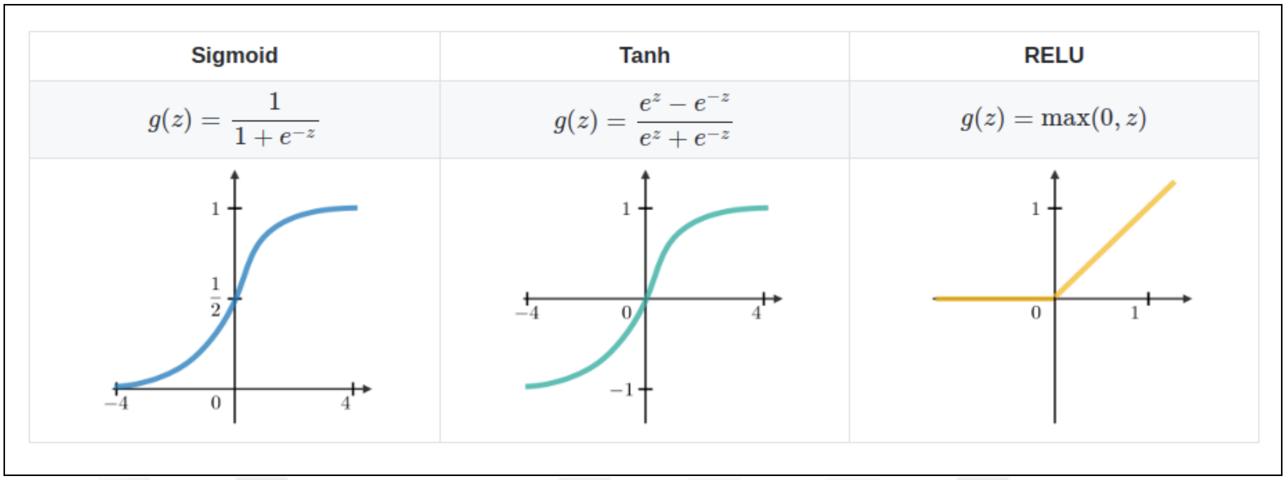

Figure 3.12: Activation function types

In the table3.3 a comparison will be made between the three types of activation functions [49] .

Table 3.3 Comparison between (Sigmoid, Tanh, Relu)

Activation Function	Range	Zero centered	Nature	Vanishing Gradient problem	Symmetric function	Model accuracy
Sigmoid	(0 ,1)	No	No linear	Yes	No	Good
Tanh	(-1,1)	Yes	No linear	Yes	Yes	Very good
Relu	(0, ∞)	No	Linear	No	No	Excellent

There are different types of artificial neural networks according to the functions of human brain cells, and there are many similarities between these types, these networks determine the outputs based on the calculations and the required set of parameters , artificial neural networks are generally divided into three main types (RNN , FNN , CNN) [50] .

a. Recurrent Neural Network (RNN): This type of network is the best to achieve results because the result return to the network, where the feedback networks feed the parameters again and the signals are transmitted in both directions, so these networks are Very powerful and complex due they have dynamic reactions that enable it to reach the equilibrium point, as it is referred to as networks recurrent and interactive , which is suitable for solving problems of optimization and correction of internal system errors e.g. competitive networks , Hopfield networks , Kohonen"s maps self-organizing[47] , the Figure 3.13 shows RNN work .

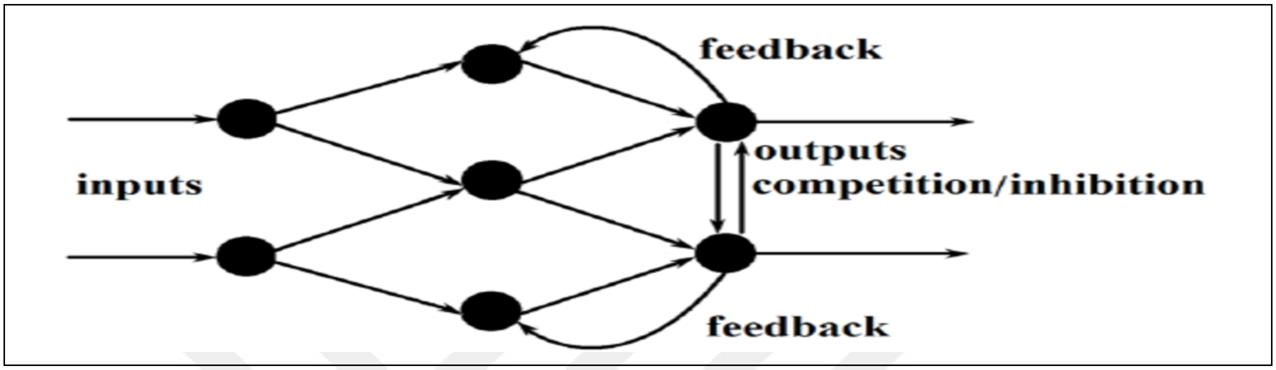

Figure 3.13: RNN work

b. Feedforward Neural Network (FNN): This type is considered basic for neural networks, where signals are transmitted in a unidirectional manner from the inputs to the processing nodes through the hidden layers (it may or not contain hidden layers)and then to the outputs, there are no loops or observations that affect the outputs in any of the layers. In addition, the inputs and outputs are fixed for this type of network, these layers are often used to classify data, e.g., Single Layer Perceptron (SLP), Multi-Layers Perceptron (MLP) as in the Figure 3.14.

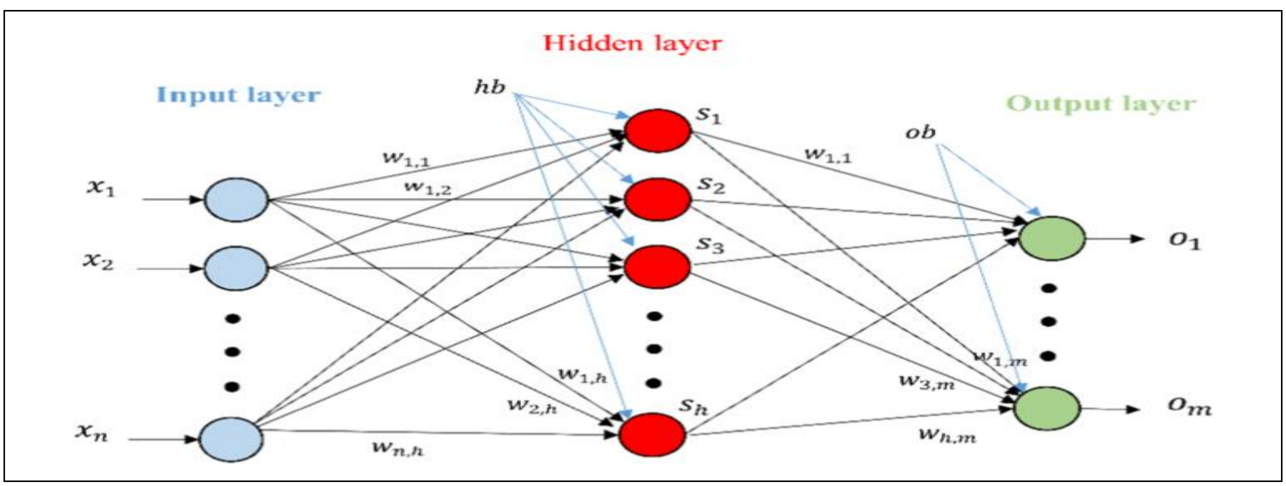

Figure 3.14: Feedforward neural network

Where x = (x1, x2, …, xn)T represents the input vector that refer n neuron values in the layer of input , s = *(s1,s2, …,sh)* T refers h output of neurons vector in hidden layer , while *o = (o1, o2,*

…, *om)*T represents m output of neurons vector in output layer , *hb* and *ob* refers biases of hidden neurons , and biases of output neurons , respectively[51] .

c. Convolutional Neural Network (CNN) : This type of algorithm is considered the most common and widely used in DL, it is a type of (FNN) where it uses convolutional structures instead of multiplying the general matrix through one layer to extract features from data without human intervention, it differs from traditional neural networks and was specifically designed for pixels data processing and also used to images recognition and processing , CNNs consist of three main layers convolution, pooling and full connected , where the convolution and pooling layers extract the features while the full connected layer maps those to the extracted features such as classification , convolution plays an important role in CNNs, where pixel values are stored in 2D array through a specialized type of linear operation , after that, a kernel is applied, which is a network of parameters that extracts an optimizer feature at each image position, making CNN more efficient and effective for process images as in Figure 3.14, [52].

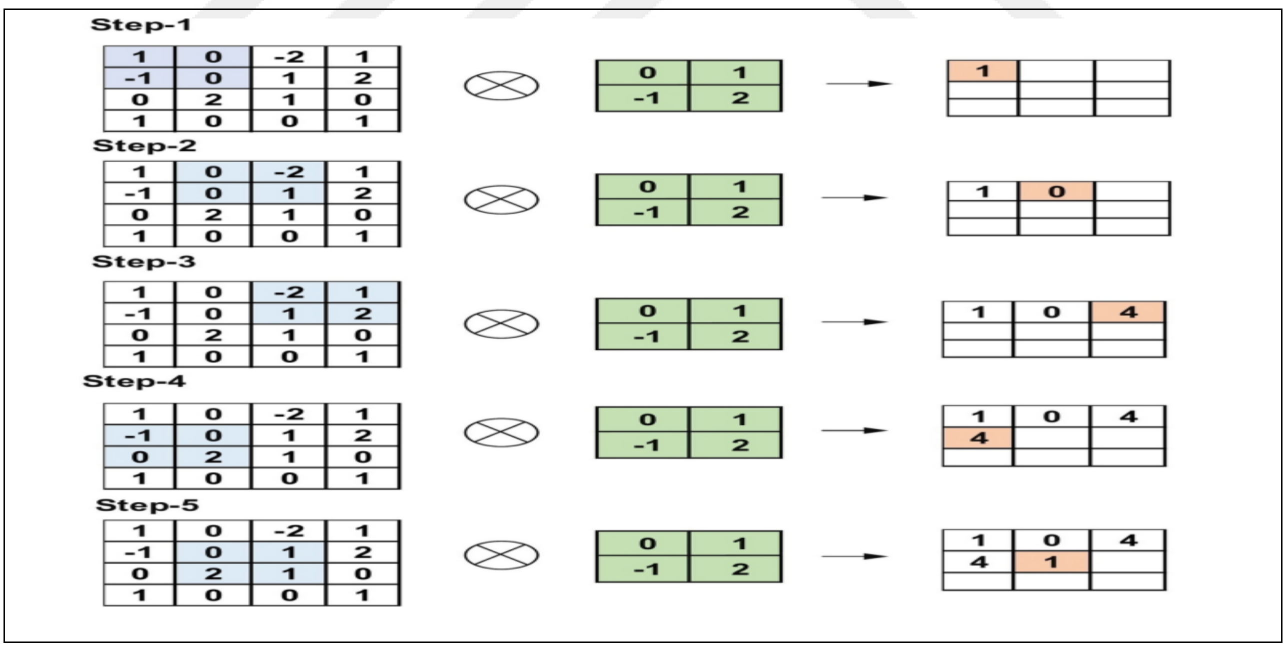

Figure 3.15: Steps of convolutional layer

The pooling layer reduces the internal dimensions of the feature maps through a typical down sampling process, so that the input is stable and small transitions and distortions are reduced, in addition to reducing the number of parameters in the later stage , the maximum pooling is the most common that extracts the maximum value and ignores other values by extracting corrections from feature maps Input and output as in Figure 3.15 [53].

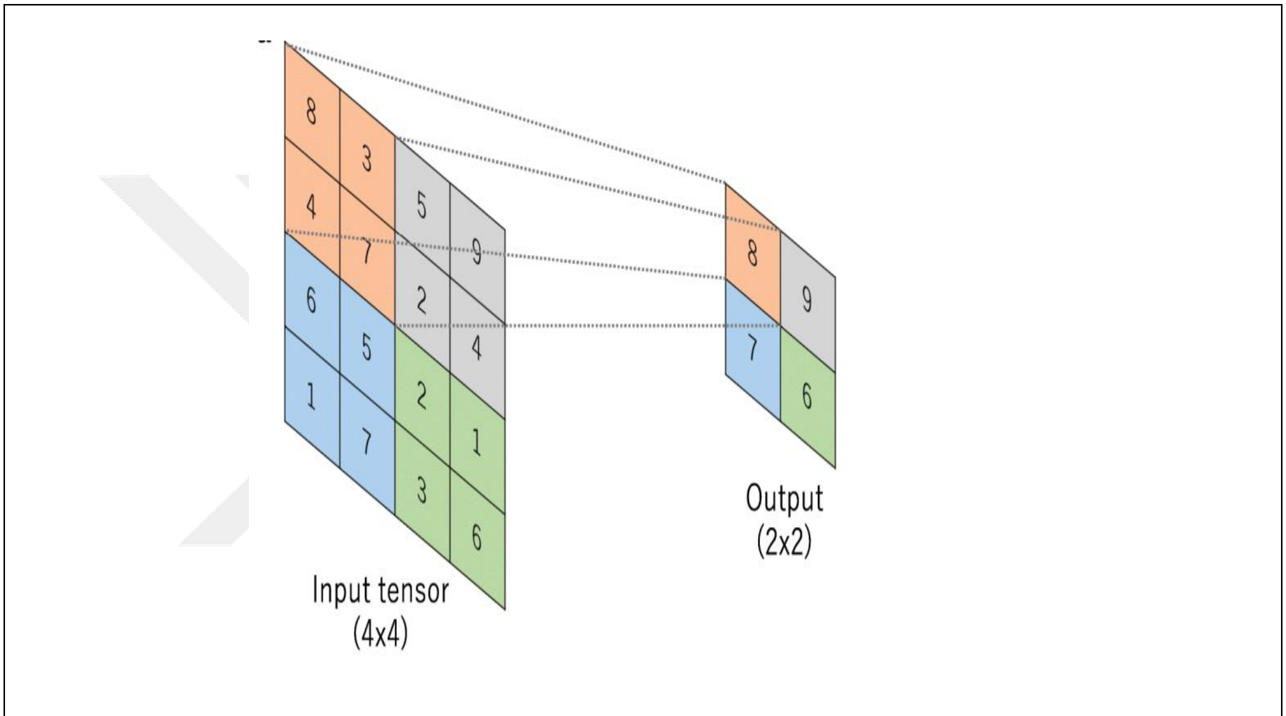

Figure 3.16: Max pooling operations

gradually extracted features can become more complex, where the function of the kernel is to optimize the parameters and this process is called training to reduce the difference between the ground truth labels and the output through a backpropagation and gradient descent algorithm , the Figure 3.16 , represents the architecture and training of 2D CNN that consist of several blocks, where the weights and kernels are calculated with the loss function through one or several fully connected layers and the input data is converted into an output through forward propagation , It is also possible with the same procedures to apply the 3D CNN forward propagation on the learnable parameters and training data set and update both weights and kernel according to the value of the loss with the rectified linear unit (Relu) through algorithm of gradient descent optimization with backpropagation [54] .

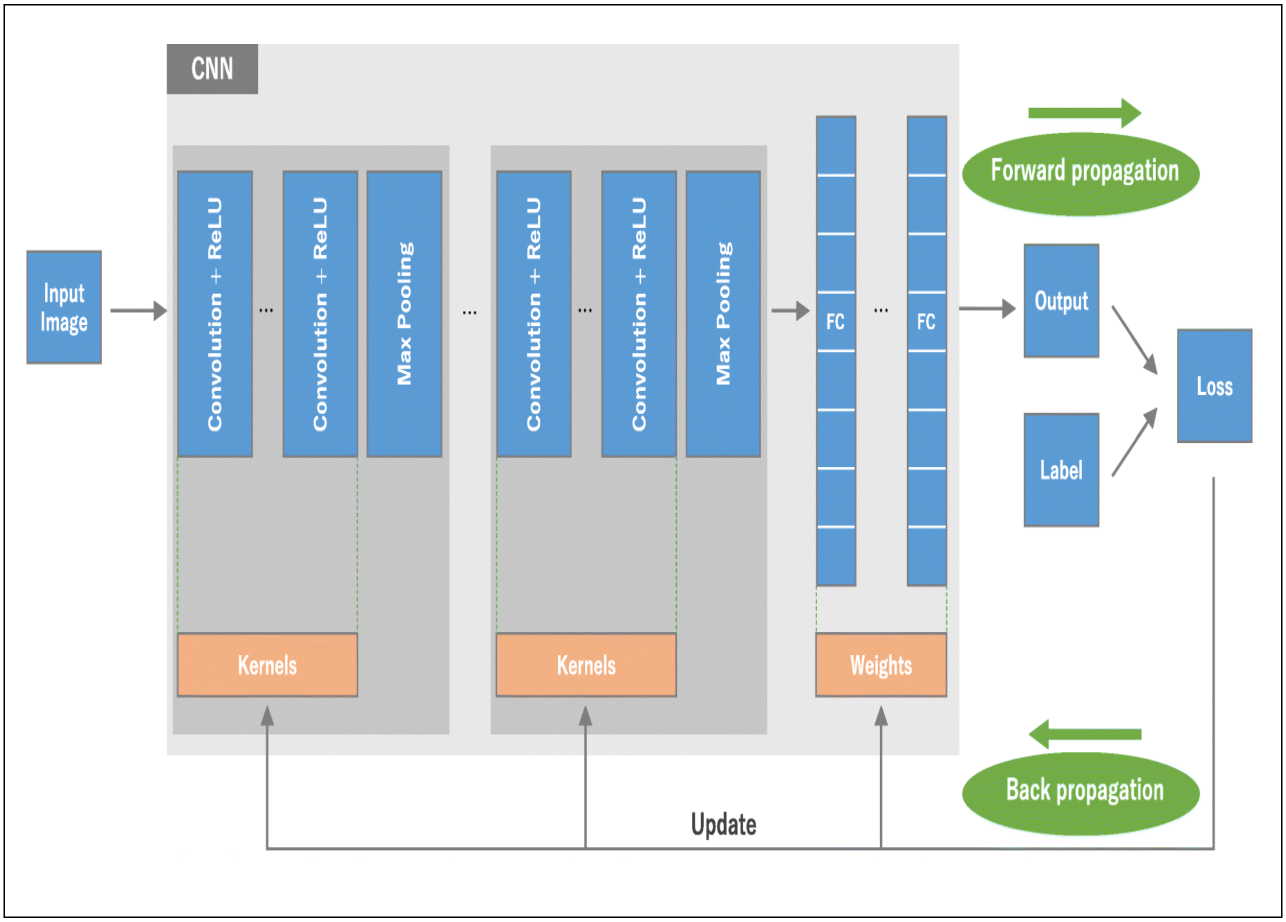

Figure 3.17: Architecture of CNN

In previous years, convolutional neural networks were applied to 2D and 3D data and able to learn patterns and complex objects through several hidden layers with the millions of parameters, where neural networks are trained on a large volume of data such as images, video, etc. in 2015, the first 1D CNN that was proposed for Serkan Kiranyaz and his colleagues deals with ECG signals and was small in size and adaptable and achieved a high classification , as it achieved high levels in many applications such as early diagnosis, classification of medical data, detection of faults in electric motors, and others, this type of network also has other advantages such as low cost for easy configuration of network compresses and implementation of devices in real time , in some applications, 1D CNN has become better than 2D CNN due to a large difference in the computational complexity as in Figure 3.17 , [55] ,[30] .

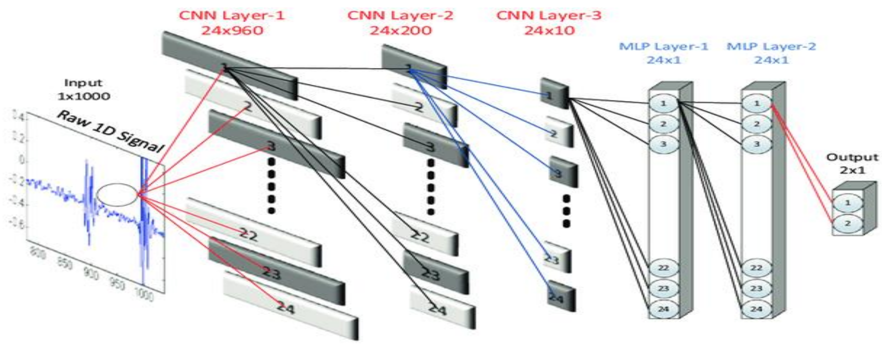

Figure 3.18: 1D-CNN configuration sample

3.2.2 Unsupervised Learning

It is an unsupervised learning style where the machine can predict the output without any supervision through the training process using an unlabeled data set and the unsupervised learning algorithms cluster unclassified data according to the input patterns in addition the similarities and differences between them , unsupervised learning is the best option in some problems when data on the desired results are not available, in addition, it is easier to obtain unclassified data than obtaining classified data [56] . In Figure 3.18 shows the method of unsupervised learning to identify the visual properties of oranges and apples through the algorithm without human intervention, as the algorithm is not trained on the input data set , unsupervised learning classified into two types [57].

a. Clustering : Objects are grouped together into groups based on similarities and differences, so the objects that have commonalities are within one group, while the objects that have few or no commonalities remain in other groups and are classified according to the common features , clustering customers according to products they purchase , the most common examples of clustering algorithms K-Means Clustering , Mean-Shift , DBSCAN .
b. Association : this type of learning refers to finding the typical relationship between variables within a large data set, such as market data analysis and mining, and it uses several algorithms Apriori, FP-Growth , Eclat .etc.

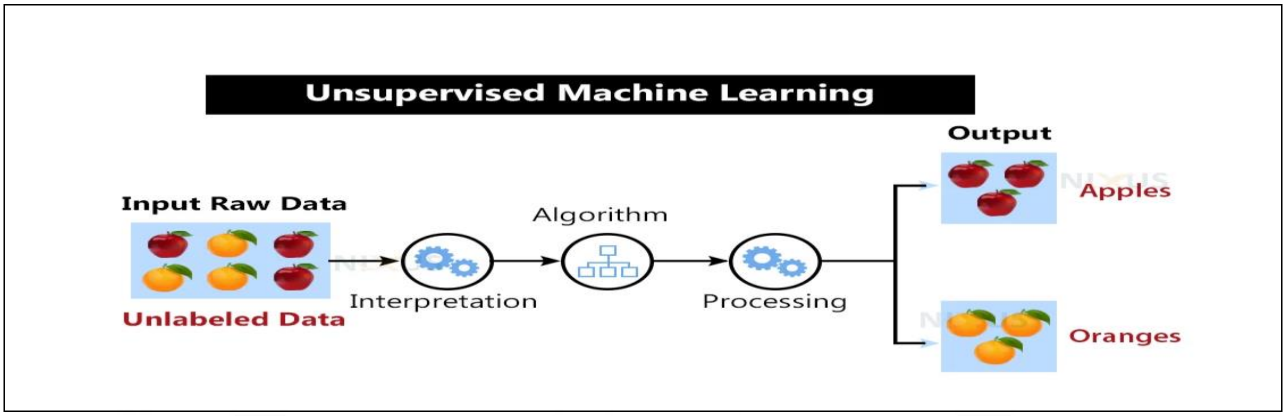

Figure 3.19: Unsupervised learning style

3.2.3 Semi-supervised Learning

This type of learning considers the characteristics of both supervised and unsupervised learning and is a mixture of unlabelled data with a labelled data set ,the defects of this type are covered by the supervised learning data , this type of learning can be applied in the medical field when collecting labelled data is very costly and time consuming, where the few available data labelled with big data is unlabelled and trained using Computer-Aided Diagnosis as in Figure 3.19 , [58] [59] .

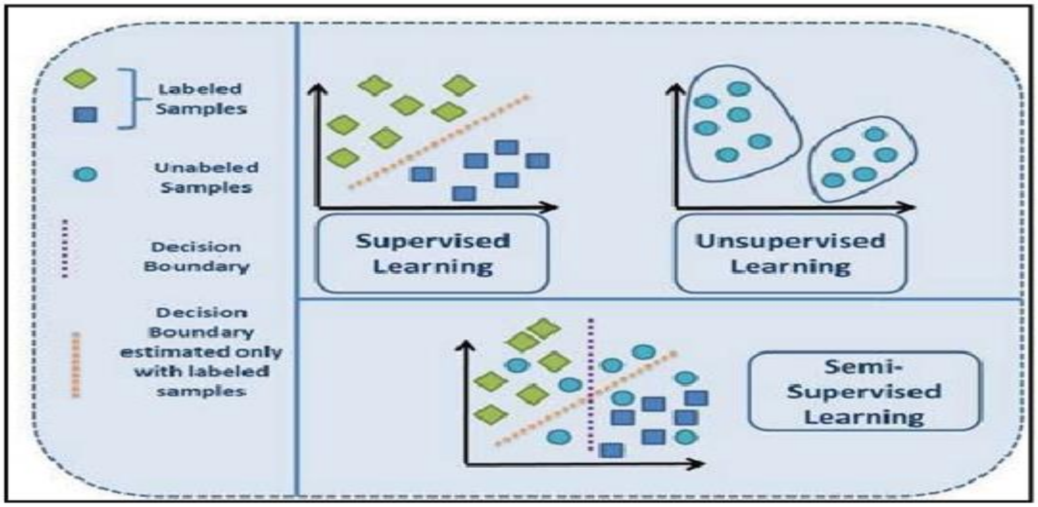

Figure 3.20: Semi-supervised learning

3.2.4 Reinforcement Learning

This type of learning is based on feedback process, where the AI component through operation and experiment evaluates its surroundings, learns from experiences and improves performance, where the component rewards every good step and punishes every bad step , this type of learning aims to maximize rewards through high performance of actions and is the unlike of subjected learning to supervise where it lacks classified data and learns from experiences, reinforcement learning is applied many various aspects as game theory, and multi-agent systems, information theory , there are two types of reinforcement learning positive and negative reinforcement learning as in Figure 3.20 , [60] .

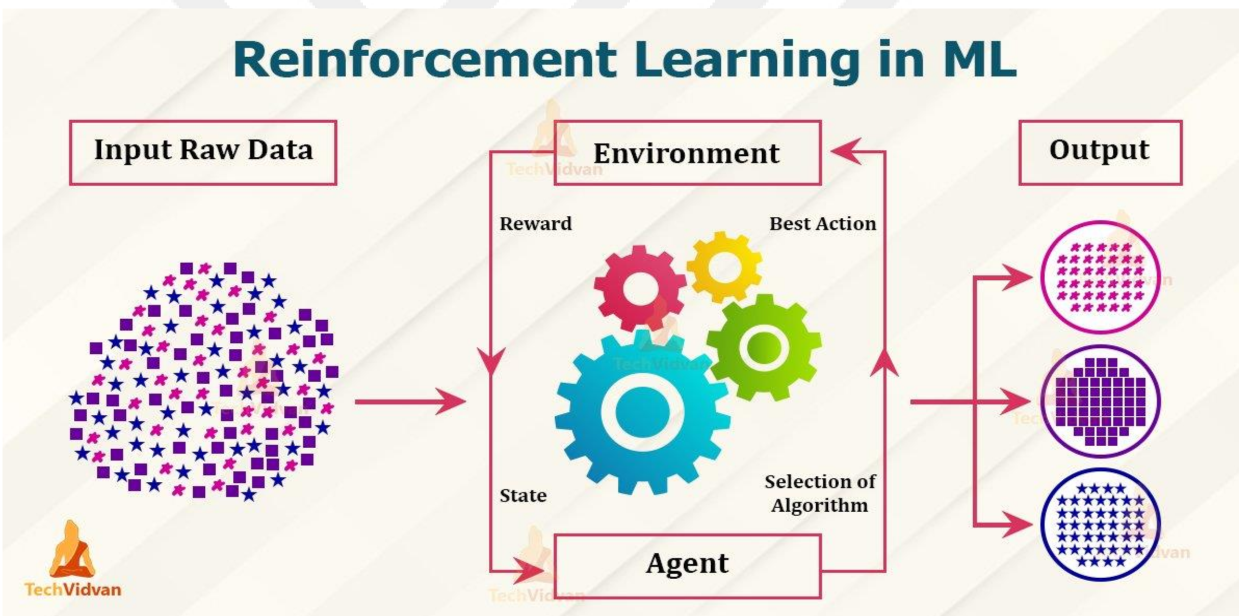

Figure 3.21: Reinforcement learning

3.3 APPLICATIONS OF ML AND DL

There are several applications of ML and DL in all fields that deal with big data around the world and most of them and the best five applications are:

a. Healthcare industry : In recent years, ML technologies have increasingly been adopted in the healthcare industry , this technology helps junior doctors analyse trends and report events that may aid in the process of improving patient prognosis and determining appropriate treatment , ML and DL also gives credit to devices that monitor patients to assess their health in real time, and the algorithms can predict diseases that may pose a threat to human life ,in addition, contribute to drug discovery and Personalized treatment In a short time, for example, the Watson system used by Pfizer IBM to discover drugs by analysing huge amounts of disparate data [61].

b. Finance sector : Many banks and financial institutions today are using DL and ML techniques and extracting basic insights from among the huge amounts of data that allow investors to decide the trading time, electronic monitoring systems, through data mining methods, focus on warning signals of fraudulent activities , for example, through many machine learning methods, pay pal is used to distinguish between fraudulent and legitimate transactions between sellers and buyers[62].

c. Retail sector : On a large scale, retail websites use ML techniques through user purchase history, where retailers use these technologies to obtain and analyse data and provide experiences that belong to their customers in addition to carrying out customer merchandise planning, marketing campaigns and price optimization [63].

d. Travel industry**:** ML plays an important role by expanding the capabilities of the travel industry. Ride services can be provided through Uber and Ola algorithms that deal with dynamic pricing ,uber uses a machine learning model called Geosurge that manages dynamic pricing parameters through predictive modelling of traffic, supply and demand patterns in real time on for example, if the meeting is late, you can get an Uber reservation immediately, but you will pay addition [64].

e. Social media : DL methods play a prominent role in the field of social media, where billions of users can communicate with each other , user-specific services can also be provided for example, in the Facebook platform, faces are recognized through automatic tagging, and the recognition of known faces begins through artificial neural network algorithms within lists Contacts[65].

4. MODELS PROPOSED AND METHODS

4.1 THE STRUCTURES OF MODELS PROPOSED

The major purpose of this study is to differentiate between acute asthma and bronchitis that affects the (LRTI) in preschool children who are less than six years old by using three algorithms of ML (KNN, DT , MLP) with two algorithms of DL (1D-CNN , LSTM) and comparing the models to get the highest accuracy and choose the most appropriate model for our study , It will work in Python (3.8.8) using Jupyter Notebook with Visual Studio Code (VSC) .

4.2 DATASET

After agreement with the consultant paediatrician , the real data was collected in Iraq at Fallujah Teaching Hospital for Women and Children and the cases were manually examined by him , where the data collection continued for four months , started in March 2022 to June 2022, we were able to collect 512 prospective cases (248 acute asthma , 264 acute bronchitis), each case contained 12 clinical features which identified by the consultant as in table 4.1.

Table 4.1: Clinical features of diseases

No	Features	Acute asthma	Acute bronchitis
1.	Sex	Mal / Female	Mal / Female
2.	Age	Under six years	Under six years
3.	Temperature	Normal	Low grade fever
4.	Runny nose	+ve /-ve	+ve
5.	Cough	+ve/ Dry	+ve/ Productive
6.	Headache	-ve	+ve
7.	Wheeze	+ve	+ve /-ve
8.	Chills	-ve	+ve
9.	Family history	+ve	-ve
10.	Shortness of Breath (SOB)	+ve	+ve
11.	General malaise	-ve	+ve
12.	Eczema	+ve	-ve

Represent (+ve) is positive and (-ve) is negative , the initial signs of the disease were set by the consultant as in the table 4.1, where there was a clear overlap in the symptoms for both diseases such as coughing, wheezing, runny nose and shortness of breath, in addition to that some cases the symptoms were not stable , as during the clinical examination they were variable according to the child’s immunity and environment [66] . The number of cases varied during the four months according to the weather fluctuations that the city experienced in that period such as (high and low temperatures, dust, air pollution, etc.) , where the cases of acute bronchitis increased in the months (March, April) when the temperatures were lower, while the cases of acute asthma increased in the months (May, June) when there were dusty weather as in table 4.2 [67][68].

Table 4.2: Number of cases distributed on months

Case type	March	April	May	June	Num	Percent %
Acute asthma	39	63	77	69	248	48.4375%
Acute bronchitis	94	78	49	43	264	51.5625 %
Total cases	133	141	126	112	512	100 %

4.2.1 Description of Dataset

After collecting the dataset of 512 prospective cases by a paediatric consultant for acute asthma and acute bronchitis, many other conditions such as pneumonia, bronchiolitis, obstructive pulmonary disease, and other diseases were excluded, and adult cases with the same disease were also excluded due the topic of our study related with children in preschool, the table 4.3 represents a sample of the collected data set.

Table 4.3: Samples of dataset

Sex	Age	Temp	Runny nose	Cough	Headache	Wheeze	Chills	Family history	SOB	General malaise	Eczema	Clinical Finding
F	5	low grade fever	+ve	productive	+ve	-ve	+ve	-ve	+ve	+ve	-ve	Acute Bronchitis
F	2	normal	+ve	dry	-ve	+ve	-ve	+ve	+ve	-ve	+ve	Acute Asthma
M	5	low grade fever	+ve	productive	+ve	+ve	-ve	-ve	+ve	+ve	-ve	Acute Bronchitis
M	4.5	normal	+ve	dry	-ve	+ve	-ve	+ve	+ve	-ve	-ve	Acute Asthma
M	4.5	normal	+ve	dry	-ve	+ve	-ve	+ve	+ve	-ve	+ve	Acute Asthma
F	5	normal	+ve	dry	-ve	+ve	-ve	+ve	+ve	-ve	+ve	Acute Asthma
F	3.5	normal	+ve	dry	-ve	+ve	-ve	+ve	+ve	-ve	+ve	Acute Asthma
M	4	low grade fever	+ve	productive	+ve	+ve	-ve	-ve	+ve	+ve	-ve	Acute Bronchitis
M	3.5	normal	+ve	dry	-ve	+ve	-ve	+ve	+ve	-ve	-ve	Acute Asthma
M	2	low grade fever	+ve	productive	+ve	+ve	-ve	-ve	+ve	+ve	-ve	Acute Bronchitis
M	5	low grade fever	+ve	productive	+ve	-ve	-ve	-ve	+ve	+ve	-ve	Acute Bronchitis
F	2	normal	+ve	dry	-ve	+ve	-ve	+ve	+ve	-ve	-ve	Acute Asthma
F	2.5	normal	+ve	dry	-ve	+ve	-ve	+ve	+ve	-ve	-ve	Acute Asthma

4.2.2 Data Pre-processing

After the collecting data by the consultant, we need to pre-process this data to make it easier to deal with by ML and DL models, as the process included converting the data into an excel sheet, taking into account accuracy during writing to be clear and easy to read, at first we verified with the paediatrician consultant by analysing the dataset and selecting the appropriate data that mainly affect the classification of diseases, in addition, some redundant data were removed, as 12 clinical features were chosen to differentiate between two diseases, and there were 6 main steps for pre-processing (as Figure 4.1)[69][70] .

a. Data profiling: We conducted a review of the data with the consultant by examining and analysing that data, identifying relevant data, and selecting important features for both diseases.
b. Data cleansing: We don’t have a missing data as it was processed by the consultant and some bad data was eliminated.
c. Data reduction: There was some excess data regarding symptoms, which were reduced and focus with the main symptoms of diseases.
d. Data conversion: Since our problem is a binary classification, the text data has been converted into numerical data (0,1) to facilitate its dealing by the proposed models as in table 4.4.
e. Data enrichment **:** To achieve a balance between the data, the proportion of data collection for both diseases was close to avoid occurrence oversampling Which may lead to the possibility of overfitting[71]. Where the proportion of acute bronchitis 51.5625 % while the proportion of acute asthma was 48.4375% as in the Figure 4.2.
f. Data validation**:** The data set was divided into two groups, the first for training the DL and ML models by 70%, while the test data was by 30% to measure the robustness and accuracy of the resulting model, the ratios were chosen in this way because our data is small and be opportunity to learn models through training is large.

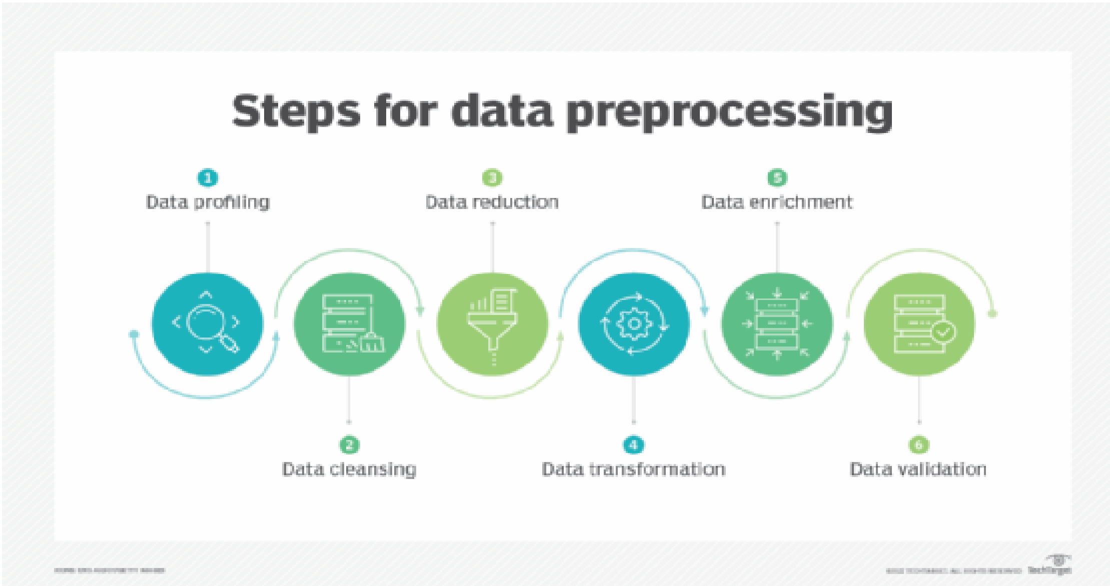

Figure 4.1: Steps for data pre-processing

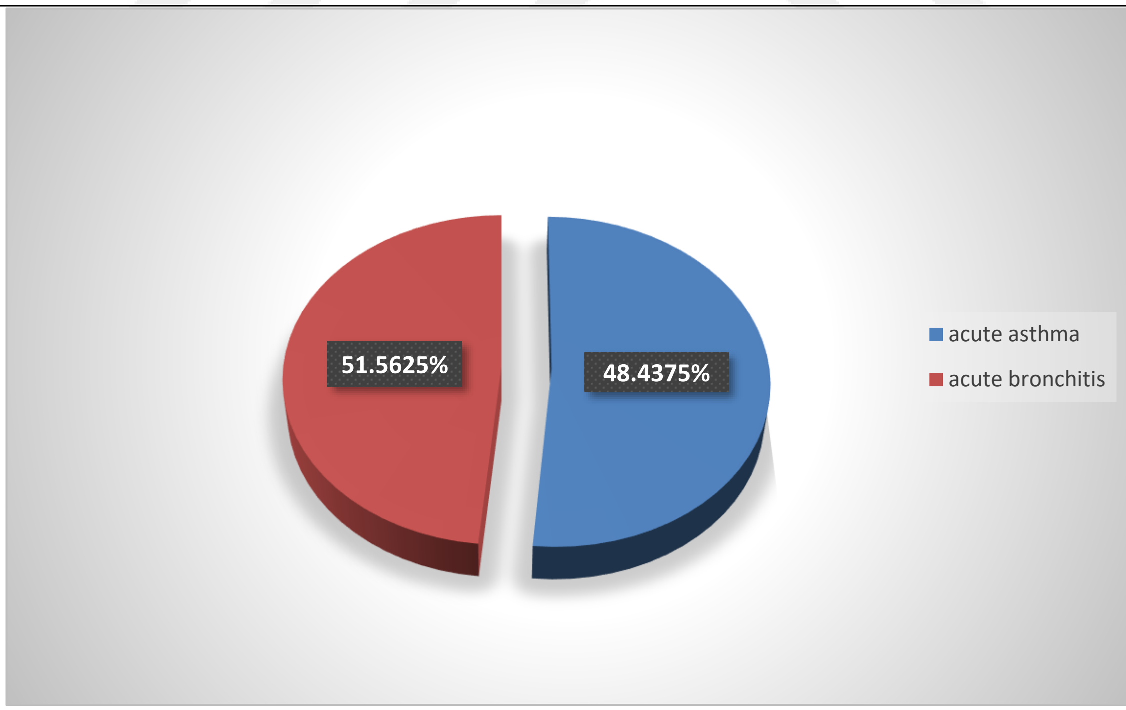

Figure 4.2: The percentage of acute asthma and bronchitis

Table 4.4: Conversion of dataset into numerical data

Sex	Age	Temp	Runny nose	Cough	Headache	Wheeze	Chills	Family history	SOB	General malaise	Eczema	Clinical Finding
0	5	1	1	1	1	0	1	0	1	1	0	1
0	2	0	1	0	0	1	0	1	1	0	1	0
1	5	1	1	1	1	1	0	0	1	1	0	1
1	4.5	0	1	0	0	1	0	1	1	0	0	0
1	4.5	0	1	0	0	1	0	1	1	0	1	0
0	5	0	1	0	0	1	0	1	1	0	1	0
0	3.5	0	1	0	0	1	0	1	1	0	1	0
1	4	1	1	1	1	1	0	0	1	1	0	1
1	3.5	0	1	0	0	1	0	1	1	0	0	0
1	2	1	1	1	1	1	0	0	1	1	0	1
1	5	1	1	1	1	0	0	0	1	1	0	1
0	2	0	1	0	0	1	0	1	1	0	0	0
0	2.5	0	1	0	0	1	0	1	1	0	0	0
1	2	0	1	0	0	1	0	1	1	0	0	0
0	2.5	1	1	1	1	1	0	0	1	1	0	1
1	5	1	1	1	1	0	0	0	1	1	0	1
0	2.5	1	1	1	1	1	0	0	1	1	0	1
0	5	1	1	1	1	1	1	0	1	1	0	1
1	4	1	1	1	1	1	0	1	1	1	0	1
1	4	1	1	1	1	0	0	0	1	1	0	1
1	3.5	1	1	1	1	0	1	0	1	1	0	1
1	5.5	1	1	1	0	0	0	0	1	1	0	1
0	3.5	1	1	1	0	0	0	0	0	1	0	1

Where female = 0, male = 1, low grade fever = 1, normal = 0, positive = 1, negative = 0, dry = 0, productive = 1, acute bronchitis = 1, acute asthma = 0 .

4.2.3 Information of Dataset

After converting the textual data set to numerical data, Python will be able to read it easily, as all data were of categorical type (0,1) except the age, where the data was continuous, the table 4.5 shows representation of the data used in python.

Table 4.5: Data representation in python

No	Attributes	Data type	Description
1	Sex	Int 64	Memory allocated for this character
2	Age	Float 64	Numeric characters with decimals
3	Temperature	Int 64	Memory allocated for this character
4	Runny nose	Int 64	Memory allocated for this character
5	Cough	Int 64	Memory allocated for this character
6	Headage	Int 64	Memory allocated for this character
7	Wheeze	Int 64	Memory allocated for this character
8	Chills	Int 64	Memory allocated for this character
9	Family history	Int 64	Memory allocated for this character
10	Shortness of Breath (SOB)	Int 64	Memory allocated for this character
11	General malaise	Int 64	Memory allocated for this character
12	Eczema	Int 64	Memory allocated for this character
13	Clinical finding	Int 64	Memory allocated for this character

4.2.4 Histogram of Dataset

The Figure 4.3 shows a histogram for each feature of the dataset with the number of categorical data, for example, the percentage of males was higher than females , in addition, the percentage of cases who had shortness of breath was higher than those who did not, and the percentage of children who had a runny nose was higher as well as those who eczema disease higher than others While the rest of the categorical data were close in terms of the percentage, the ages of the children were different, and the highest percentage was recorded for the two-year-old.

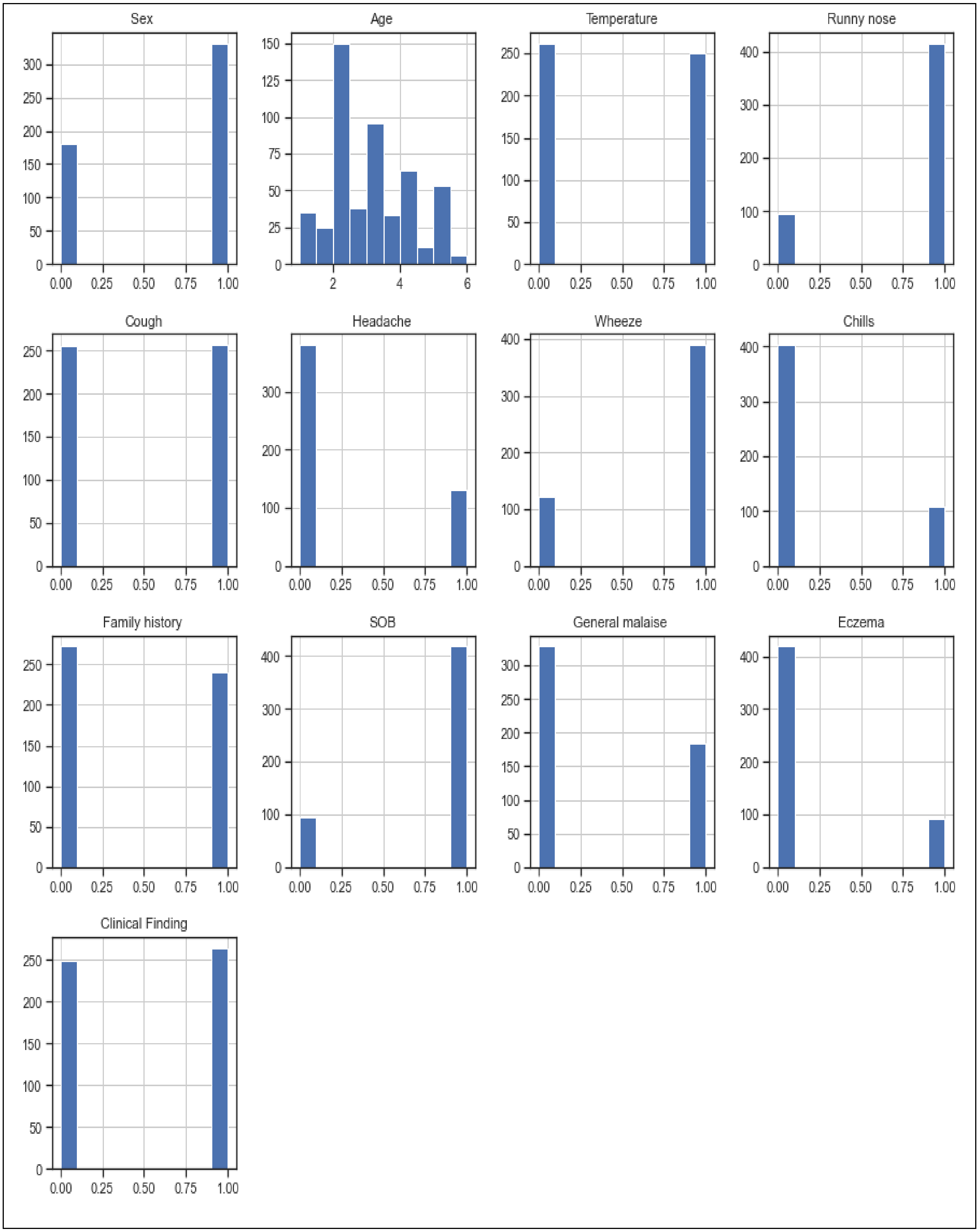

Figure 4.3: Histogram of dataset

4.2.5 Standardization of Dataset

Standardization (Z) is an important step in data processing, as the data for our study will be unified and converted into one common format although all our data is categorical except for age, it is continuous, in which the mean (μ) and standard deviation (σ) are will be 0,1 respectively as in equations (4.1,4.2,4.3), in order for the data to become more clear during processing, in addition to the ease during its analysis and processing by ML and not necessary to applied in decision tree, because it is not sensitive to the feature scale, as it divides the nodes based on one feature that increases the homogeneity of the nodes and this division not affected by other features , and might be benefit in decision tree if compare performance with other of methods [72][73].

$$standardization \qquad z = \frac{x-\mu}{\sigma} \tag{4.1}$$

$$with\ mean \qquad \mu = \frac{1}{N}\sum_{i=1}^{n}(x_i) \tag{4.2}$$

$$Standard\ deviation \qquad \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(x_i-\mu)^2} \tag{4.3}$$

4.2.6 Correlation Matrix

Correlation is a statistical measure to represent the relationship between two dataset variables, and there are two types of variables First, when the relationship between the two variables is positive, the two variables move in the same direction, but when the two variables move in opposite , the relationship is negative, Correlation is a cause and effect relationship between variables through hypothesis testing and is also used to predict trends in the real world , Since the correlation does not mean causation in some cases, so the relationship can be causal if a relationship positive strong , Where there are two ways to calculate the relationship between the two variables, the first is the Pearson correlation coefficient, which is used to measure the strength of the linear relationship between two variables, it ranges from -1 to 1 , Calculation the correlation coefficient r between two variables X and Y as in equation 4.4. The second is the Spearman correlation coefficient to measure the strength and direction of the relationship between two variables, and it ranges from -1.0 to +1.0. It is used when the two variables are not normally distributed and is considered better than Pearson's coefficient because it can be determined whether the relationship is linear or not between two variables as Pearson's coefficient is also known as the correlation coefficient between rank variables as in equation 4.5[74] .

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2}} \qquad (4.4)$$

Where *x* , *y* bar is the mean value of *x* and y ,while *xi* and *yi* refer different values of *x* and *y*.

$$r_s = 1 - \frac{6\, \Sigma\, d_i^2}{n(n^2 - 1)} \qquad (4.5)$$

Where *n* is the number of pairs for dataset, while *di* is the difference between the two ranks

The Figure 4.4 represents the correlation matrix between variables for the data set used in our study, where the Seabron library was used within Python to visualize the data to determine the most correlated pairs among the variables, Through the heatmaps of the correlation, the strength of the relationship between the numerical variables of the data set used in our study can be found, as follows:

a. A strong positive correlation between the variables (cough, temperature), (cough, general malaise) , (temperature and general malaise) ,this can be explained by the existence of a multiple linear relationship between the variables because the Pearson correlation coefficient is greater than 0.7.

b. A strong negative correlation between the variables (temperature, family history) , (cough , family history) .

c. A strong positive correlation between some variables and target (clinical finding) such as temperature, cough, general malaise, where they effect on the target was higher than 0.7.

d. A strong negative correlation between family history and target (clinical finding).

The analysis of the data set through the correlation matrix is very useful for doctors which the physician can find the relationship between two features and make appropriate medical decisions. It also helps in reducing the dimensions for large medical data by dropping one of the features with a strong positive relationship to prevent repetition that may affect overfitting and thus affect the performance of the final model[75][76].

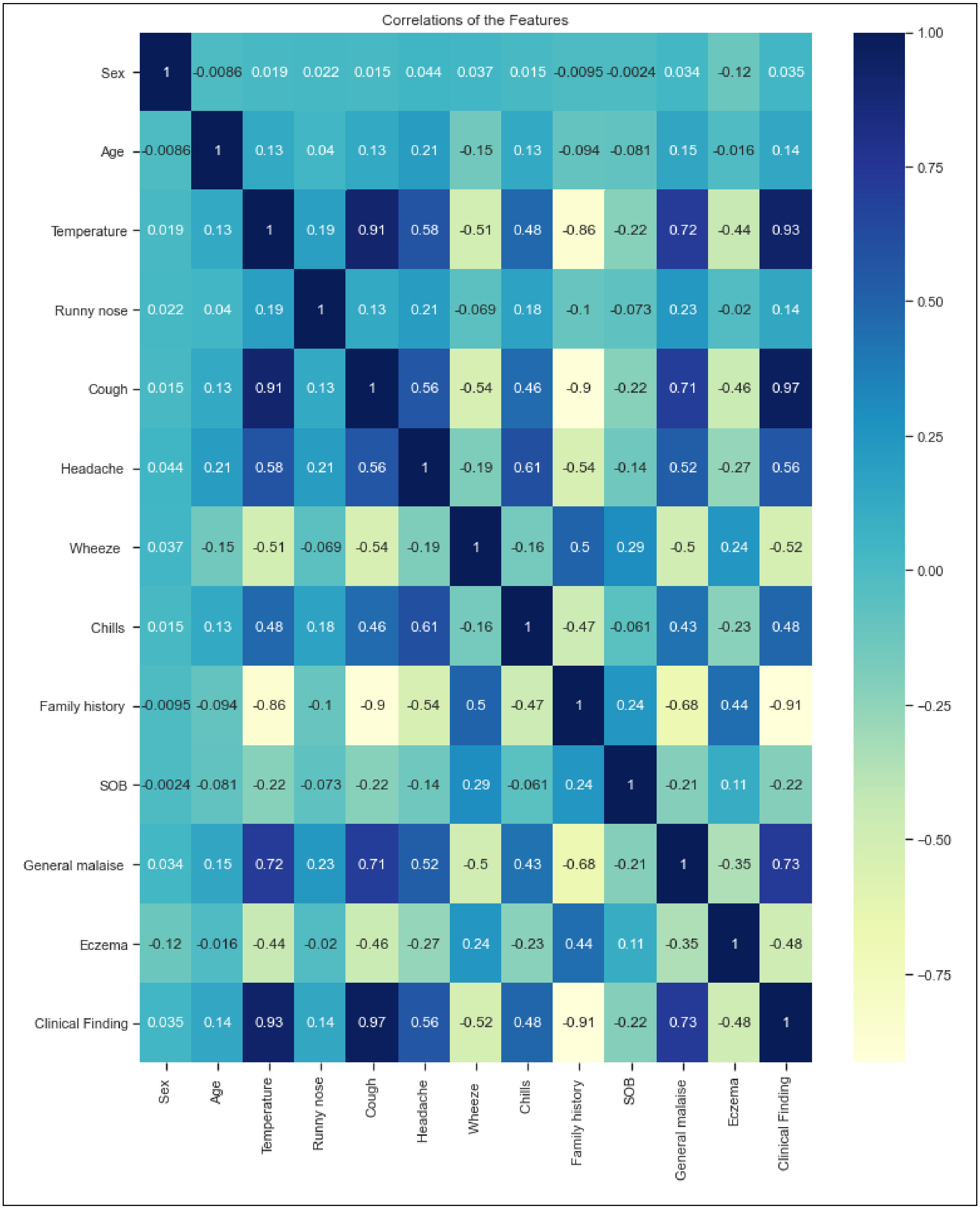

Figure 4.4: Correlation matrix of dataset

4.3 THE METHODS OF MACHINE LEARNING

By reviewing the previous literature of ML models and their applications in the field of health care, specifically within the classification problems of diseases, and after reviewing the results that they achieved and overcoming many of the problems they encountered, three models of machine learning selected (K-NN , DT, MLP),In addition, they are more suitable with the size and type of data for this study [2][17][22]. The figure 4.5 represents building framework of ML.

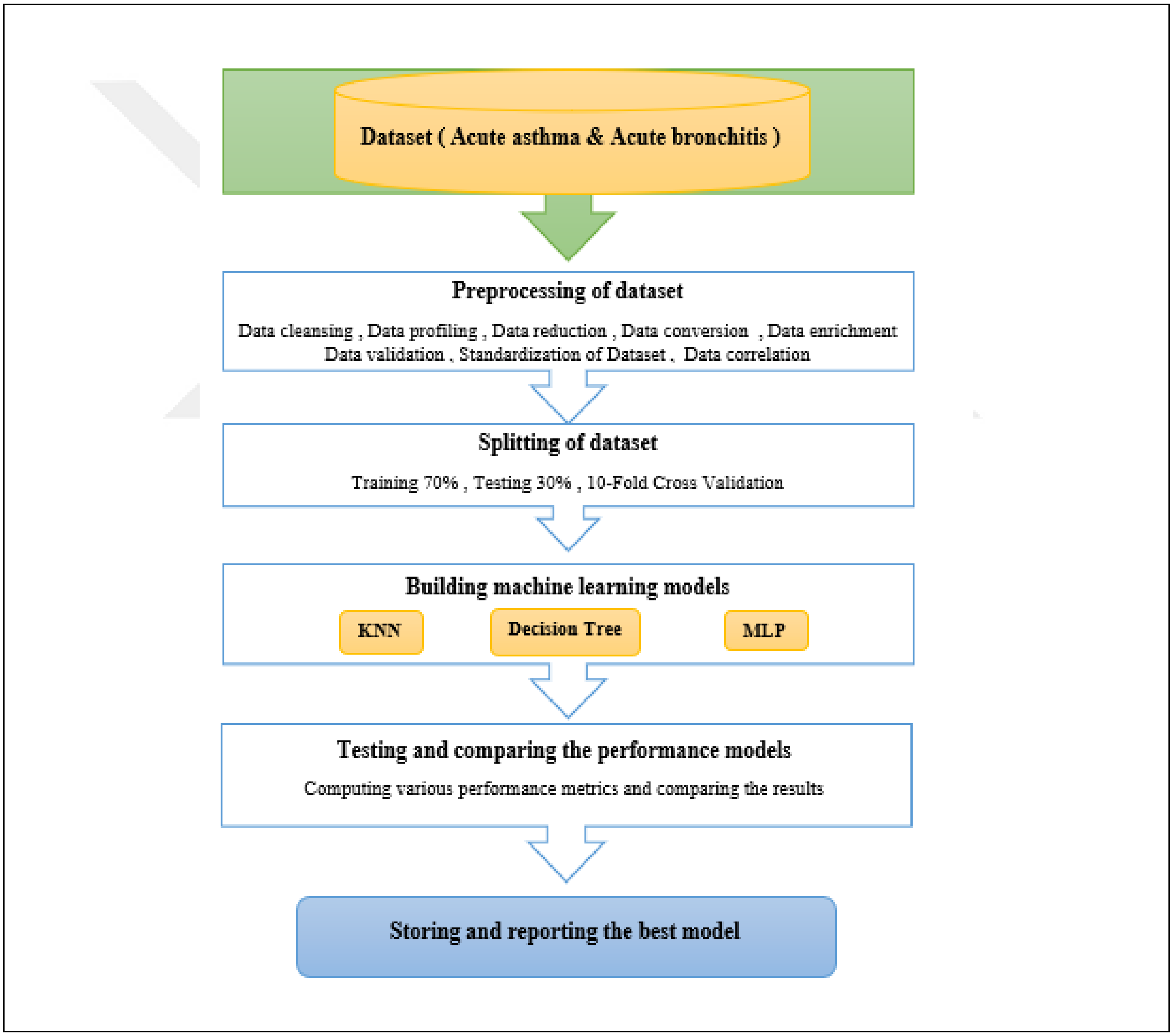

Figure 4.5: Building a framework of machine learning

4.3.1 Cross Validation Technique

After the data processing is completed and to ensure that the models are working accurately and in order to obtain superior performance of the acute asthma and bronchitis data set that was divided into 70% training data and 30% test data , the cross-validation technique implemented through the use of (GridSearchCV) within the scikit-learn library in Python, in the cross-validation process the training data will be divided into two parts, the first part is training data and the second part is validation where k- 10 fold cross validation was used for our dataset, which is an iterative process that divides the training data into (k) sections and keep in each iteration one section for the test and the (k-1) other sections to train the model as in figure 4.6 ,during each iteration, it will give us the performance of the model and in the end it gives an average of the overall performance and the process takes a long time to evaluate the best hyperparameters such as (estimator, parameter grid search , scoring, cross validation , ROC_AUC , early stopping)[77][78].

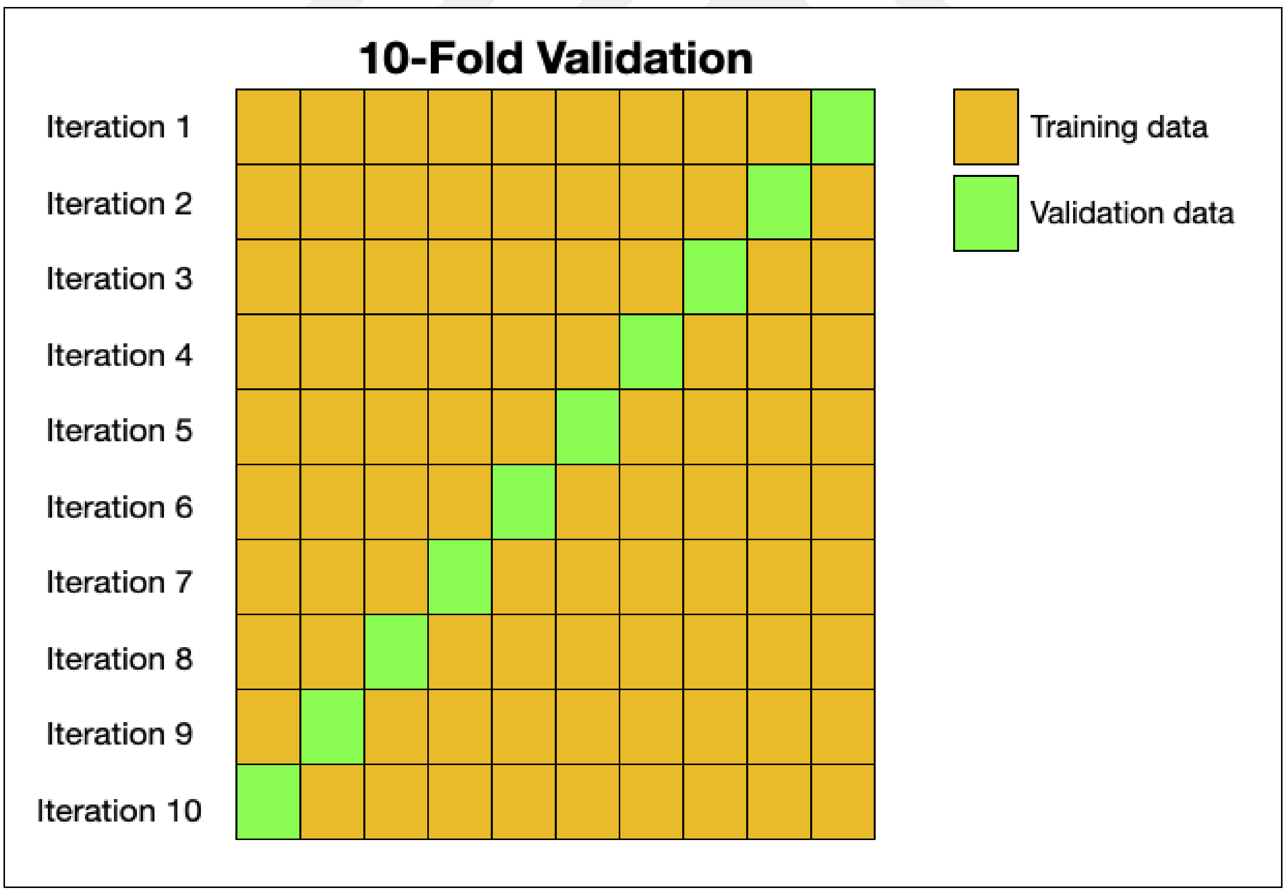

Figure 4.6: 10- Fold cross validation

4.3.2 AUC - ROC Curve

AUC – ROC curve is considered one of the most important measures to verify the performance of any classification model, and since the problem of our study is classification, AUC (Area Under The Curve) represents the measure or degree of separability it was applied to the data set to tell us about the model's ability to distinguish between categories , where the higher the AUC is at predicting as 0 and 1, the better the performance of the model to differentiate between acute asthma (green distribution curve) and bronchitis (red distribution curve) as in figure 4.7 , where AUC it is close to 1, the result is good, while if it is close to zero, the result is bad and If it is 0.5, it does not have the ability to separate the categories .

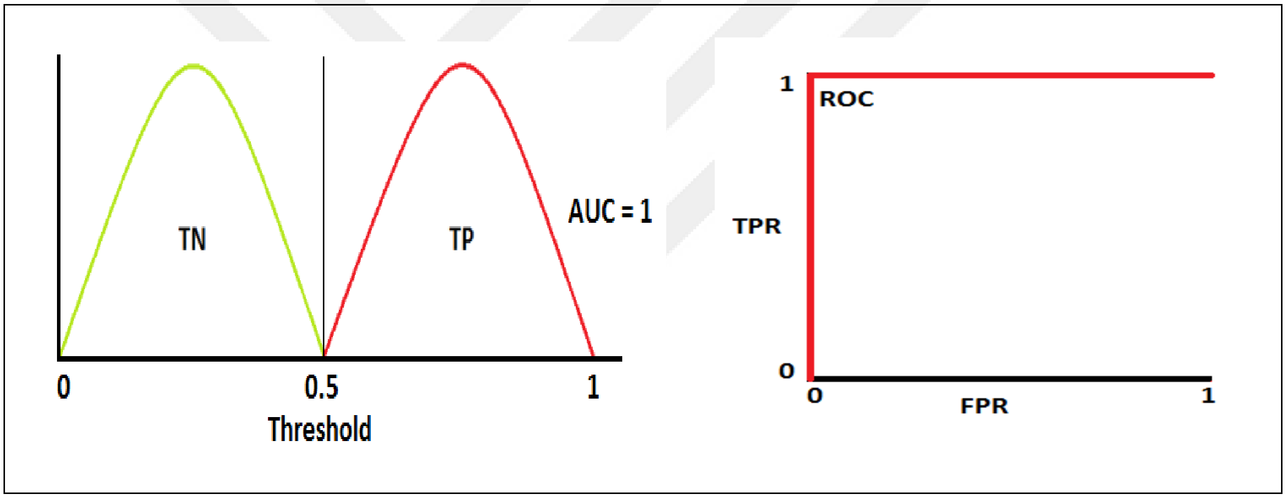

Figure 4.7: AUC - ROC Curve

ROC (receiver operating characteristic curve) is a curve of probability for a graph explain the classification model performance at all classification thresholds, the curve has two parameters: True Positive Rate (TPR) and False Positive Rate (FPR) they can be represented as in equations 4.6 , 4.7 [79].

$$TPR = \frac{TP}{TP + FN} \qquad (4.6)$$

$$FPR = \frac{FP}{FP + FN} \qquad (4.7)$$

4.3.3 Building of Classification Models

4.3.3.1 Build of k-nearest neighbours

After determining the best hyperparameters to differentiate between acute asthma and bronchitis, a nearest neighbour model is built , the main idea is to this algorithm based on discovering the data points closest to K, which are known as the neighbours of the predicted data point, where 5 nearest neighbours of the data will be selected and calculated by Euclidean distance , Additionally, the parameters k (n- neighbours) are optimized to obtain the best parameter by creating a list of possible k parameters from (1 to 31) , fitting 10 fold for each of 30 candidate to be totalling 300 fit [80] .

4.3.3.2 Build of decision tree

The decision tree for the data set of our study will be built by selecting the best trait from among the data set for the root nodes as well as the sub nodes using the Attribute Selection Measure technique (ASM) for overall features The supported criteria were also used as Gini index and entropy, and selection of the maximum depth of the tree from (1 to 20) , fitting 10 fold for each of 40 candidate to be totalling 400 fit [81] .

4.3.3.3 Build of multilayer perceptron

The multilayer Perceptron will be built using the input layers for the dataset, passing through the hidden layers and then the output layers to get a binary classification, in the hidden layers multi layers are set as (5,5,5) , (5) , (6), (4) , It will choose the best hidden layer that gives better results, as well as set batch size between (auto, 100) in addition to the max iteration (500, 600) with learning rate (0.001,0.005, 0.01) , and through the hyperparameters, the best results will be selected [22].

4.3.3.4 Confusion matrix

The confusion matrix or error matrix is a common metric for solving binary and multi-class classification problems in which the performance of an algorithm is evaluated through a matrix where each row represents the cases existing in actual class while each column represents the cases existing in a predicted class, the Figure 4.8 show the confusion matrix to the binary classification for our dataset.

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

Figure 4.8: Confusion matrix to the binary classification

Where actual 0 represented acute asthma and actual 1 represented acute bronchitis and TN refers to the true negative which explain the number examples negative classified similarly TP refers to the true positive which explain the number examples positive classified while FP refers number of examples actual negative classified as positive, and FN represented number of examples actual positive classified as negative , through this matrix, the accuracy of the classification model is measured as in the equation 4.8

$$Accuracy = \frac{TN + TP}{TN + FP + FN + TP} \qquad (4.8)$$

as the specificity and sensitivity has been applied for our data set which mathematically represented the accuracy of a test which refers the presence or absence of a case as in equations 4.9 and 4.10.

$$Sensitivity = \frac{TP}{TP + FN} \qquad (4.9)$$

Where sensitivity indicating the possibility of a positive test (true positive rate) if it truly being positive.

$$Specificity = \frac{TN}{TN + FP} \qquad (4.10)$$

Where specificity indicating the possibility of a negative test (true negative rate) if it truly being positive , and applied the precision (also called value of positive predictive) is the part of relevant cases among the retrieved cases among the retrieved cases, while the recall (also refer as sensitivity) is the part of relevant cases that were retrieved therefore , both recall and precision are based on relevance as in equations 4.11 and 4.12 , However, there is need to used F-Score for evaluate performance through metric which of required to predict of the performance as in equation 4.13 [82][21] .

$$Precision = \frac{TP}{TP + FP} \tag{4.11}$$

$$Recall = \frac{TP}{TP + FN} \tag{4.12}$$

$$F - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4.13}$$

4.4 THE METHODS OF DEEP LEARNING

By reviewing the previous literature of deep learning models and their applications in the field of health care, specifically within the classification problems of diseases, and after reviewing the results that they achieved and overcoming many of the problems they encountered, two models of DL selected (1D-CNN , LSTM) in addition, they are more suitable with the size and type of data for this study [29] [30] [35] .The Figure 4.9 represents building framework of DL.

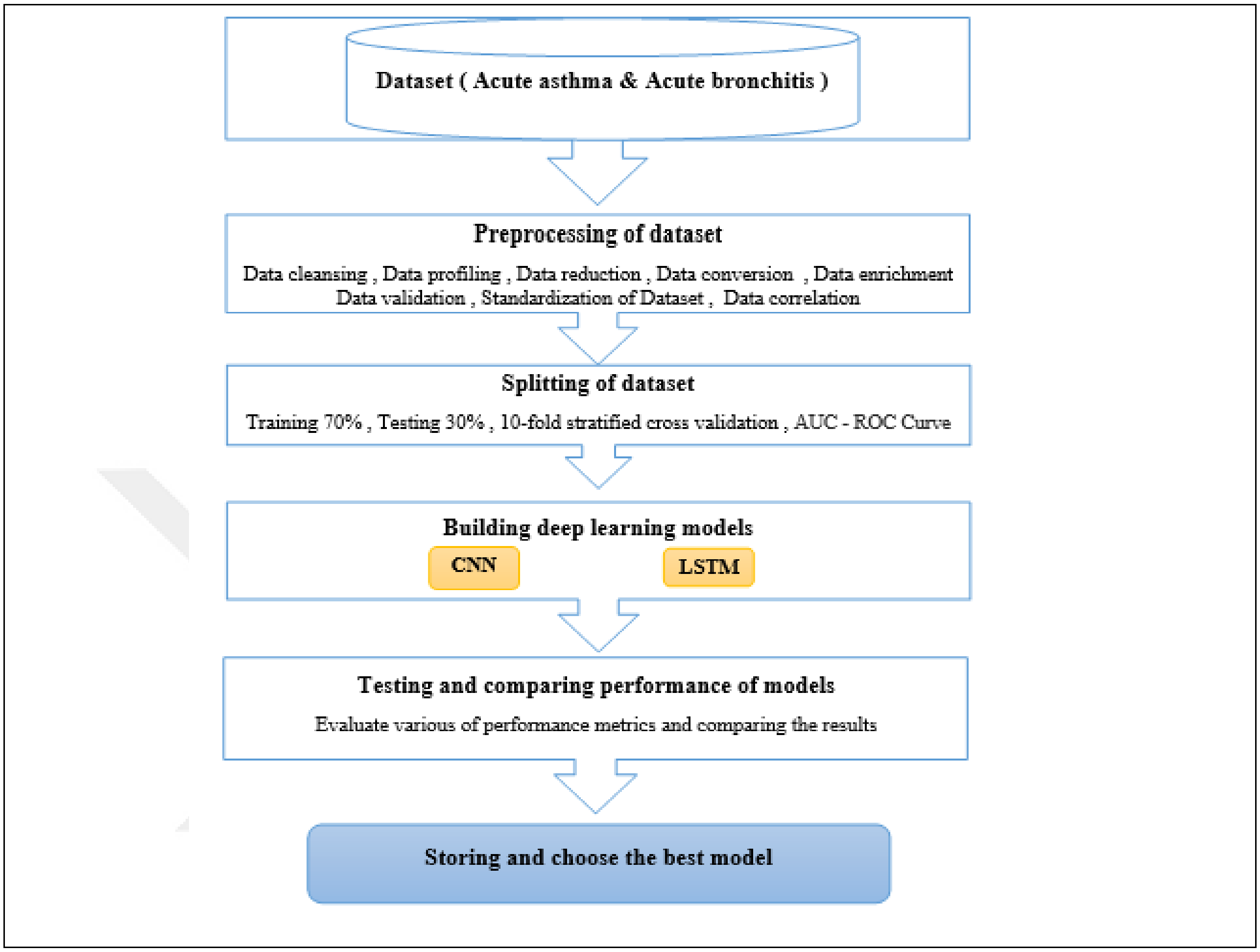

Figure 4.9: Building a framework of DL models

4.4.1 Building of Classification Models

4.4.1.1 Build of one - dimensional convolutional neural network

due our data is text (numeric) , 1D - CNN model will be built as in Figure 4.6 , where the input layer will be a 1D array (12,1) representing the input data set after it has been split into 70% training as array (358,12) and 30% testing as array (154, 12) , and we will use three convolutional layers 1D with filters (32, 64, 128) respectively with the use of a kernel size of (3) for each convolutional layer to extract the features from the matrix and improve the output after it is multiplied by the input as in Figure 4.7 in addition used activation function (Relu) with each layer.

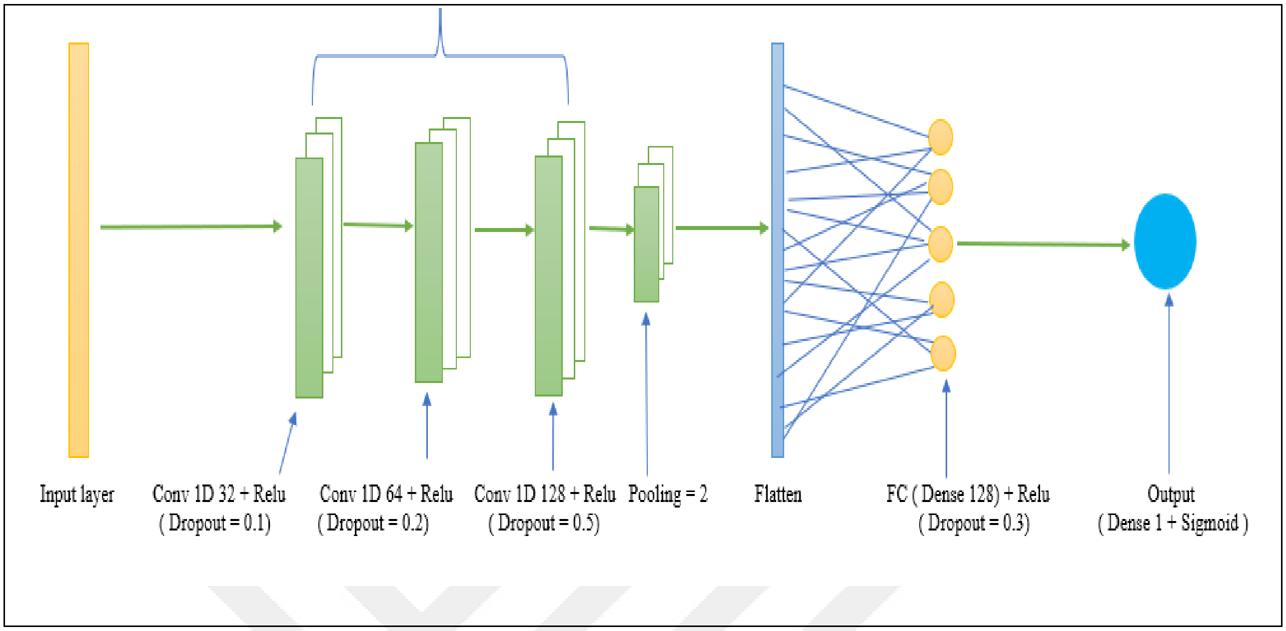

Figure 4.10: Building of 1D - CNN model

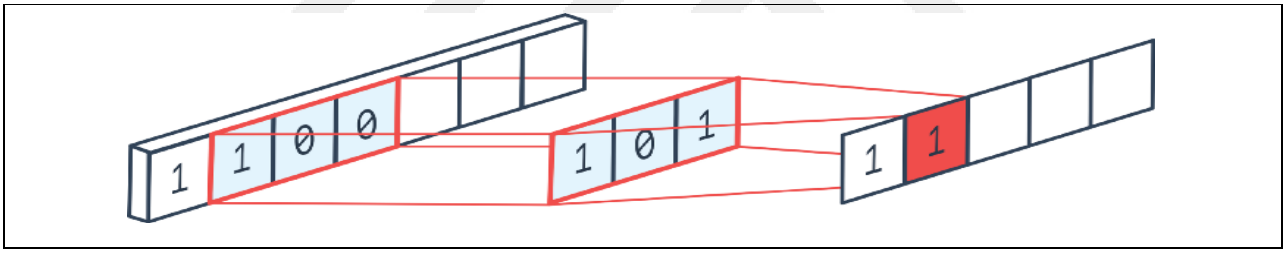

Figure 4.11: Single channel of 1D convolution with 3 kernel size

Where the output will be one-dimensional because the filter moves only in one direction, and to avoid overfitting, layers were dropout as rate (0.1,0.2, 0.5) after the three convolutional layers respectively [83] . To reduce the size of the data and the number of parameters, a max pooling layer of size 2 was added to extract the features while reducing the variance by keeping the largest value in the pooling rectangle as in Figure 4.8.

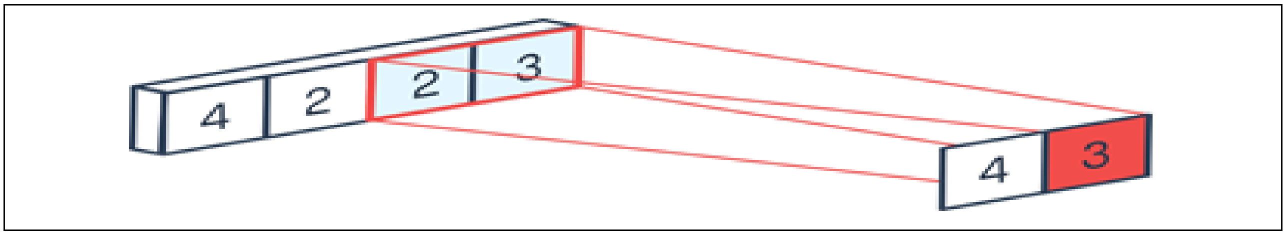

Figure 4.12: Max pooling of 1D_ CNN

Then the outputs of the previous layers are normalized by adding a flattening layer and making it into one vector and preparing it for the input in the next stage, and to adjust the weights, a fully connected layer was added with a filter 128 and a dropout of 0.3, and add the final layer for outputs with activation function sigmoid to get the final classification of either acute asthma or acute bronchitis , and the table 4.6 show 1D- CNN configuration for our dataset .

Table 4.6 :1D - CNN configuration

Layers	Filters	Kernel size	Activation function	Dropout
Conv 1D	32	3	Relu	0.1
Conv 1D	64	3	Relu	0.2
Conv 1D	128	3	Relu	0.5
Max pool 1D = 2	-	-	-	-
Fully connected	128	-	Relu	0.3
Fully connected	1	-	Sigmoid	-

4.4.1.2 Build of long short-term memory

The second model of deep learning that has been built is LSTM , which is an advanced version of the Recurrent Neural Network (RNN) standard that suffers from the vanishing gradient problem because it contains short-term memory, especially with long sequential data , therefore, LSTM can insert large blocks of memory and move information forward and keep it from previous sequential parts instead of connected hidden units, thus can solving the vanishing scaling problem[34]. The LSTM model of our text dataset was built by where the input layer will be a 1D array (12,1) representing the input data set after it has been split into 70% training as array (358,12) and 30% testing as array (154, 12), where two equal layers of LSTM will be created, each layer contains 100 memory units of smart neurons with the addition of return sequences after the first layer that re-outputs the hidden state at the time of input for each step , with dropout after each layer of LSTM at a (0.2) percent to reduce overfitting , after building LSTM layers, the fully connected layers were added in respectively , in addition to the use of the Relu activation function with each layer and the dropout in different proportions according to the size of the layer and the end used the sigmoid activation function to obtain the binary classification , the table 4.7 show configuration of LSTM and the Figure 4.9 show building of LSTM .

Table 4.7: Configuration of LSTM

Layers	Memory unit	Activation function	Dropout
LSTM	100	-	0.2
LSTM	100	Relu	0.2
Fully connected	128	Relu	0.25
Fully connected	64	Relu	0.2
Fully connected	32	Relu	0.1
Fully connected	1	Sigmoid	-

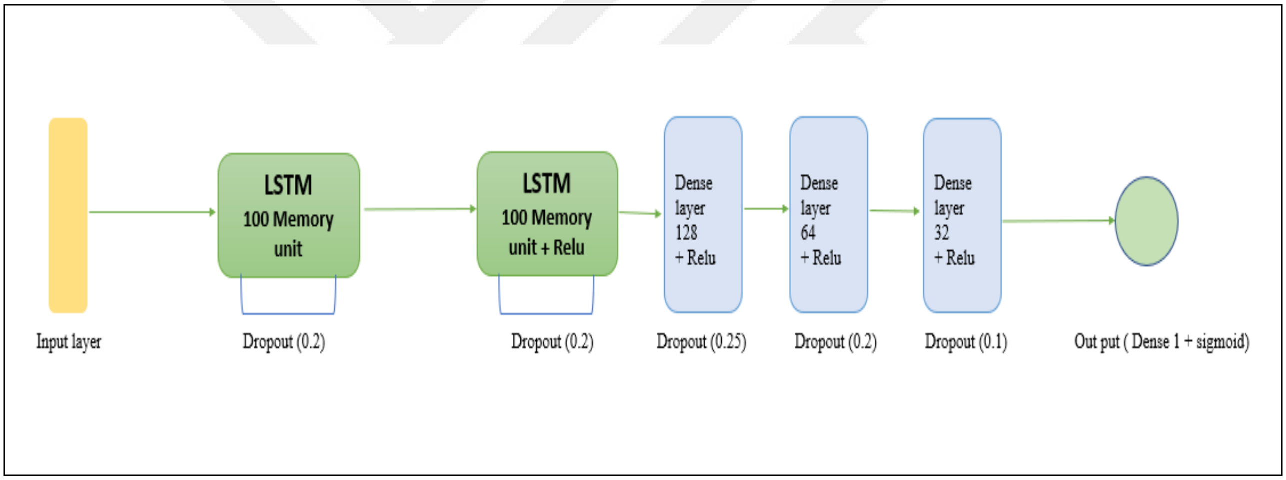

Figure 4.13: Building of LSTM model

5. RESULTS AND DISCUSSION, MODEL SELECTION

In this chapter, the results of ML through the models used will be presented and compared, then the results of DL will be presented and compared, the results will be discussed, and then the model that obtained the highest accuracy will be selected for the data set of acute asthma and bronchitis for the total of 512 real cases.

5.1 RESULTS OF ML MODELS

5.1.1 Results of K- NN

After training the K_NN algorithm for the dataset of our study, the results were obtained according to table (5.1), (5.2).

Table 5.1: Techniques used in K- NN

K	K _ range (1,31)	Distance	Technique type	CV	Split data set
5	1	Euclidean	GridSearchCV	10-Fold	70% training 30% testing

Table 5.2: Results of K_NN performance

Test case (Acute asthma)	Test case (Acute bronchitis)	Sensitivity	Specificity	ROC- AUC	Accuracy
75	79	1.0	0.9620	0.9801	98.0519

After testing the model and obtaining the final accuracy, the confusion matrix is calculated for the data set   as in table 5.3 and Figure 5.1.

Table 5.3: Results of K_NN confusion matrix

Class	Precision	Recall	F1-score	Support cases
Acute asthma	0.96	1.0	0.98	75
Acute bronchitis	1.0	0.96	0.98	79
Macro avg	0.98	0.98	0.98	154
Weight avg	0.98	0.98	0.98	154

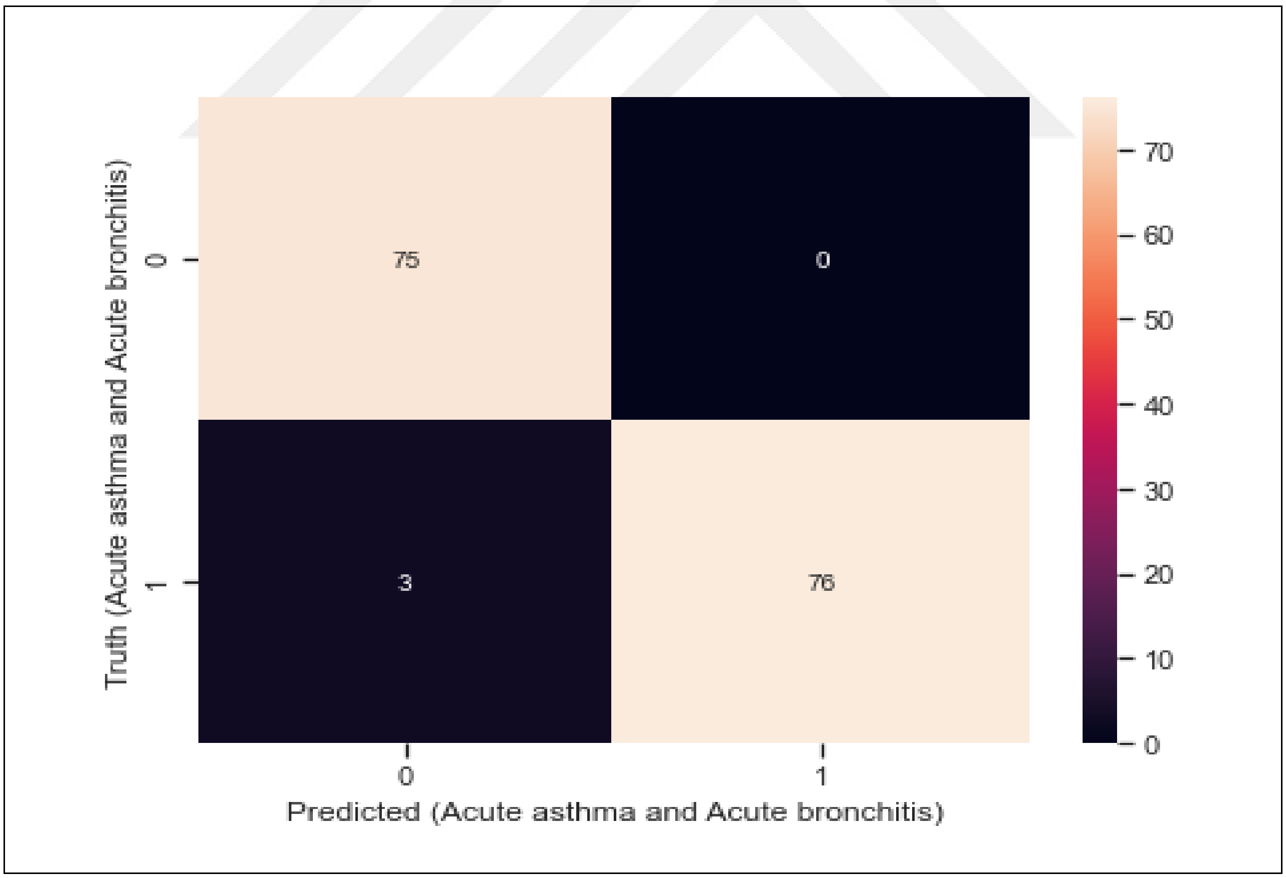

Figure 5.1: Confusion matrix of K-NN model

5.1.2 Results of DT

After training the decision tree algorithm for the data set of acute asthma and acute bronchitis, the results were obtained according to table (5.4), (5.5).

Table 5.4: Techniques used in decision tree

Max depth (1,20)	Criterion (Gini , Entropy)	Technique type	CV	Split data set
1	Gini	GridSearchCV	10-Fold	70% training 30% testing

Table 5.5: Results of decision tree performance

Test case (Acute asthma)	Test case (Acute bronchitis)	Sensitivity	Specificity	ROC- AUC	Accuracy
75	79	1.0	0.9367	0.9944	96.7532

After testing the model and obtaining the final accuracy, the confusion matrix is calculated for the data set as in table 5.6 and Figure 5.2.

Table 5.6: Results of decision tree confusion matrix

Class	Precision	Recall	F1-score	Support cases
Acute asthma	0.94	1.0	0.97	75
Acute bronchitis	1.0	0.94	0.97	79
Macro avg	0.97	0.97	0.97	154
Weight avg	0.97	0.97	0.97	154

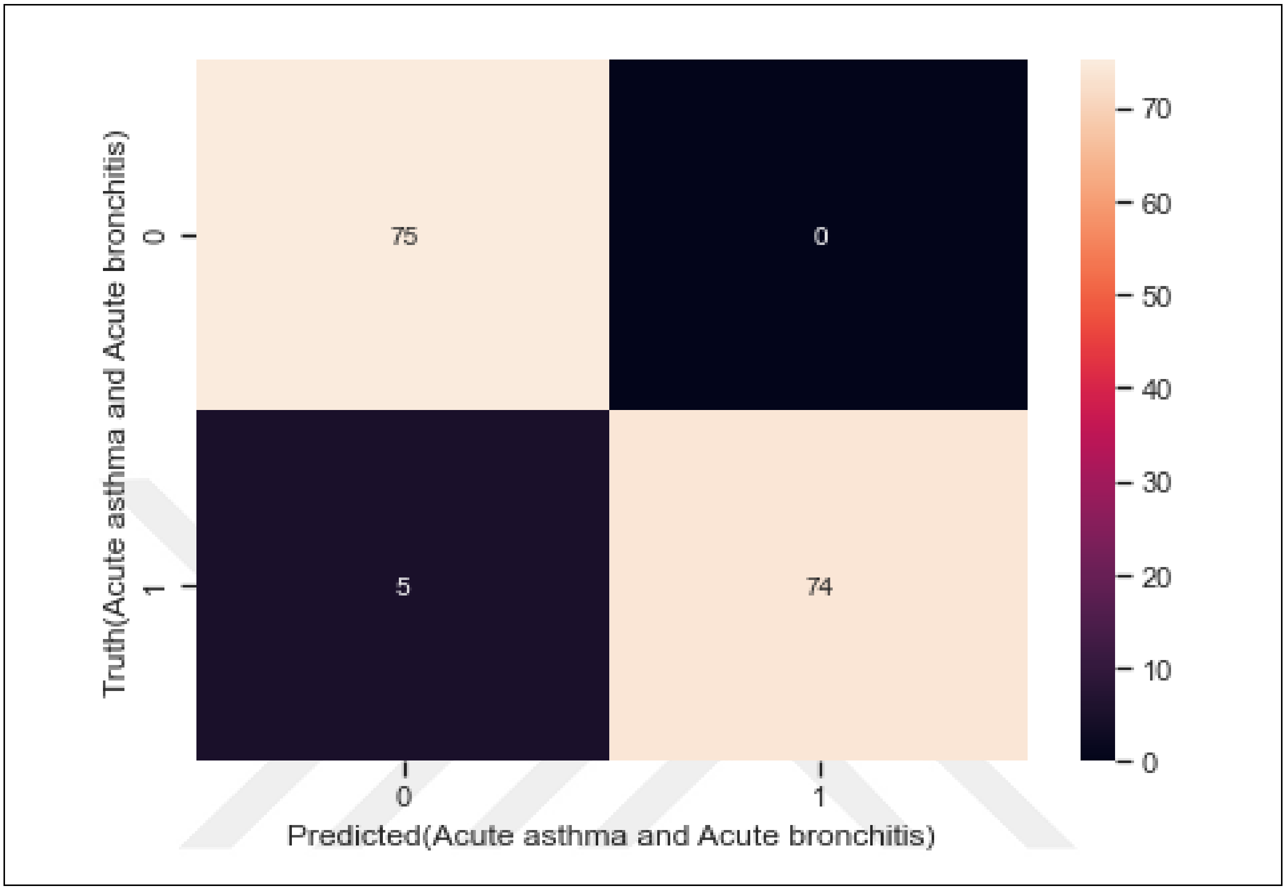

Figure 5.2: Confusion matrix of decision tree model

5.1.3 Results of MLP

After training the MLP algorithm for the data set of acute asthma and acute bronchitis, the results were obtained after the testing according to table (5.7), (5.8).

Table 5.7: Techniques used in MLP

Batch size (Auto , 100)	Hidden layer (5,5,5) (5) , (6) , (4)	Learning rate (0.001,0.005,0.01)	Technique type	CV	Max _iteration (500,600)
Auto	5	0.001	GridSearchCV	10-Fold	500

Table 5.8: Results of MLP performance

Test case (Acute asthma)	Test case (Acute bronchitis)	Sensitivity	Specificity	ROC- AUC	Accuracy
75	79	1.0	0.9746	0.1	98.7012

After testing the model and obtaining the final accuracy, the confusion matrix is calculated for the dataset as in table 5.9 and Figure 5.3.

Table 5.9: Results of MLP confusion matrix

Class	Precision	Recall	F1-score	Support cases
Acute asthma	0.97	1.0	0.99	75
Acute bronchitis	1.0	0.97	0.99	79
Macro avg	0.99	0.99	0.99	154
Weight avg	0.99	0.99	0.99	154

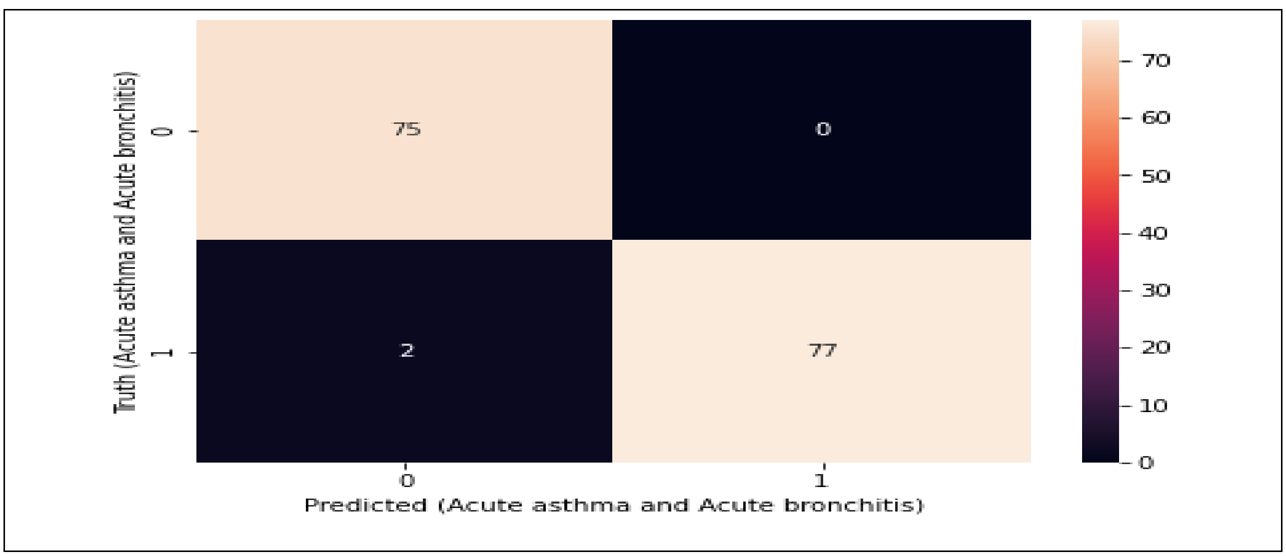

Figure 5.3: Confusion matrix of MLP model

5.1.4 Comparing Between ML Results

After the appearance of the results of the ML models and their comparison, it was found that MLP outperformed the rest of the models with an accuracy of 98.7012 to differentiate between acute asthma and bronchitis, as in table 5.10.

Table 5.10: Comparison of ML Results

Classifier	Split data set	CV	ROC- AUC	Sensitivity	Specificity	F1-score	Accuracy
MLP	70% training 30% testing	10-Fold	0.1	1.0	0.9746	0.99	98.7012
K_NN	70% training 30% testing	10-Fold	0.9801	1.0	0.9620	0.98	98.0519
DT	70% training 30% testing	10-Fold	0.9944	1.0	0.9367	0.97	96.7532

5.2 RESULTS OF DL MODELS

5.2.1 Results of 1D-CNN

The convolutional neural networks were trained using deep learning techniques, and the results appeared as in the tables (5.11), (5.12), (5.13).

Table 5.11: Techniques used in 1D-CNN

Split data	Technique type	CV	Optimizer	Loss	Epoch	Early stopping
70% training 30% testing	Stratified sampling	10	Adam	Binary cross entropy	50	Monitor = Val _loss Mini delta = 0 Mode = auto Patience = 20

Table 5.12: Results of 1D-CNN performance

Test case Acute asthma	Test case Acute bronchitis	Sensitivity	Specificity	ROC- AUC	Test loss	Accuracy	Time
75	79	0.9866	1.0	0.9932	0.0432	99.3506	12 ms

Table 5.13: Number of parameters used in 1D-CNN

Total parameters	Trainable parameters	Non trainable parameters
80499	80499	0

The results of test accuracy and test loss for 1D-CNN model appeared as in Figure 5.4 and 5.5

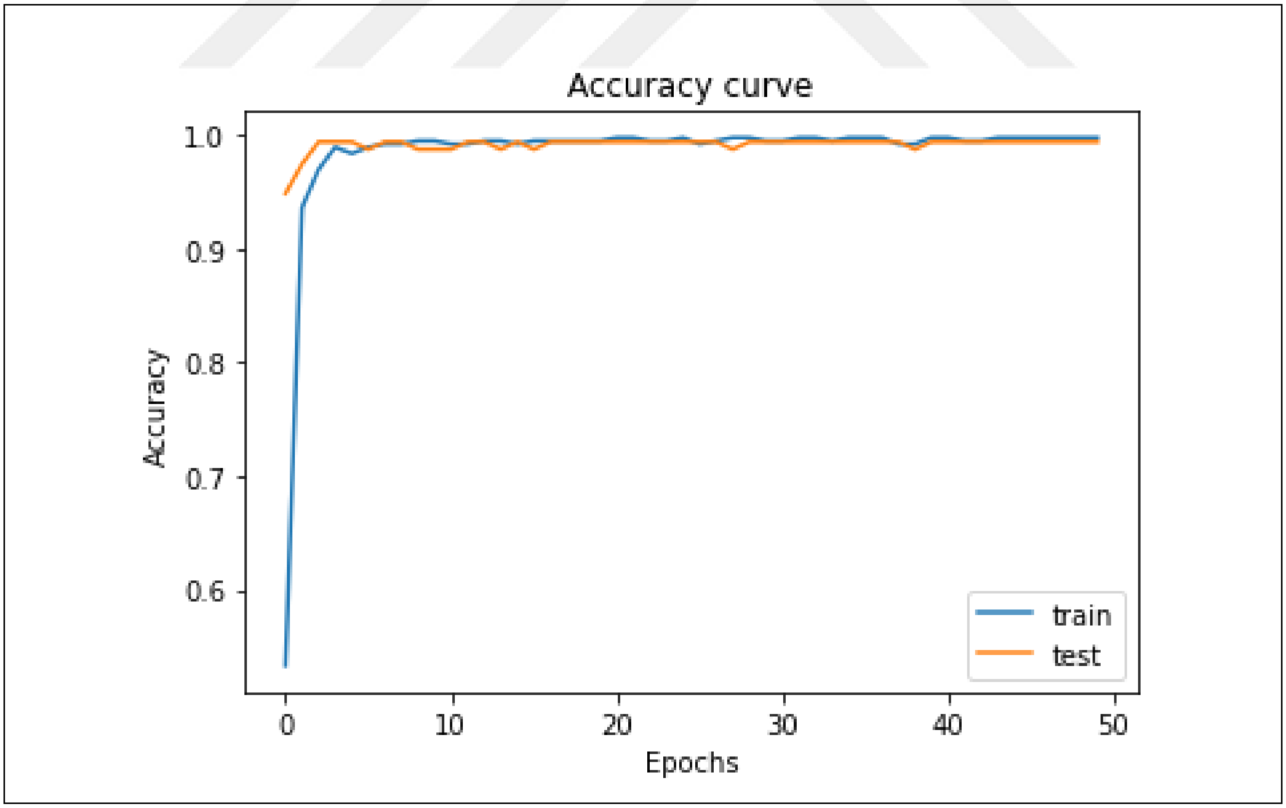

Figure 5.4: Accuracy curve of 1D-CNN model

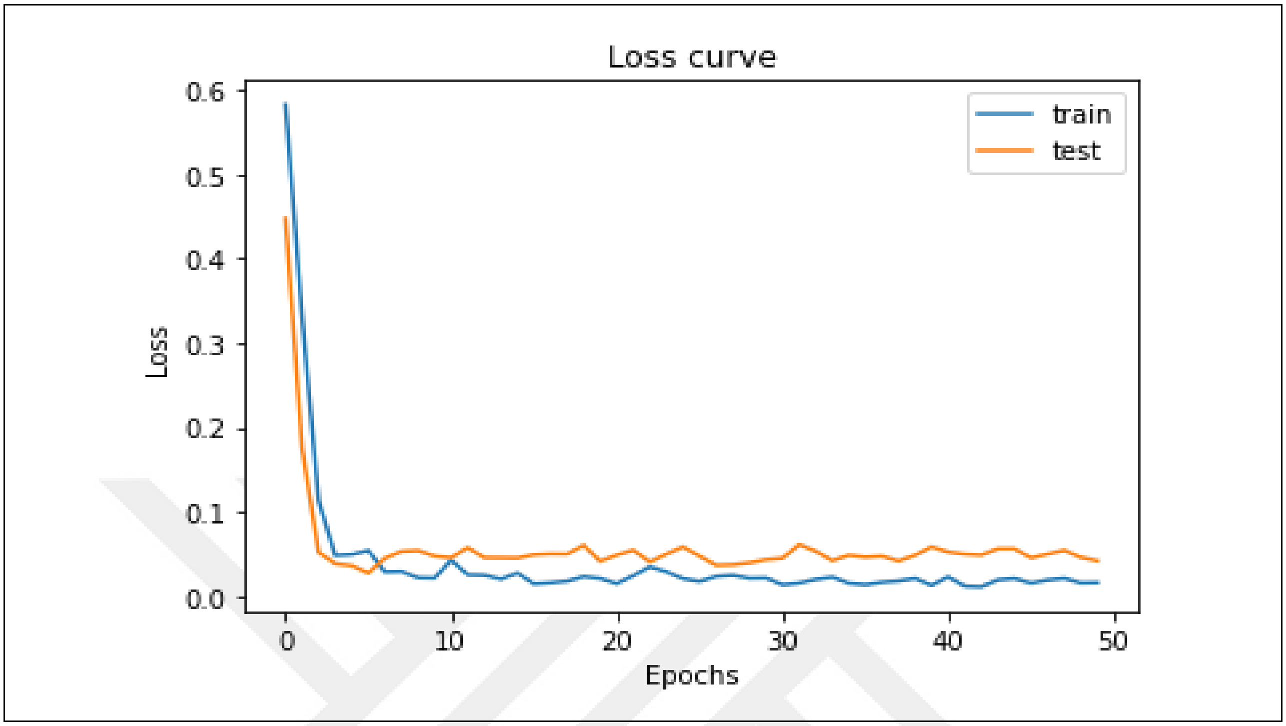

Figure 5.5: Loss curve of 1D-CNN model

After testing the model and obtaining the final accuracy with test loss, the confusion matrix is calculated for the dataset as in table 5.14 and Figure 5.6.

Table 5.14: Results of 1D-CNN confusion matrix

Class	Precision	Recall	F1-score	Support cases
Acute asthma	1.0	0.99	0.99	75
Acute bronchitis	0.99	1.0	0.99	79
Macro avg	0.99	0.99	0.99	154
Weight avg	0.99	0.99	0.99	154

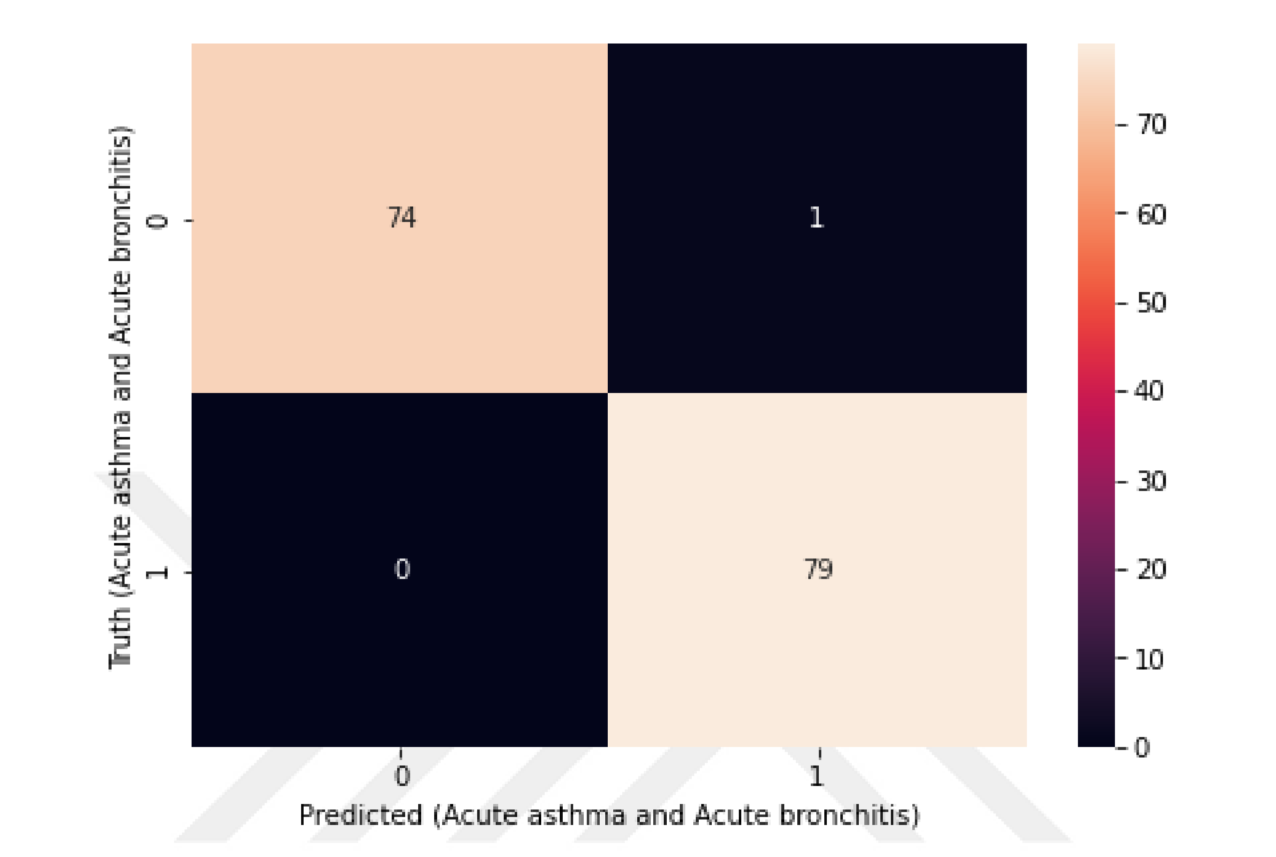

Figure 5.6: Confusion matrix of 1D-CNN model

5.2.2 Results of LSTM

The long short-term memory was trained using DL techniques, and the results appeared as in the tables (5.15), (5.16), (5.17) .

Table 5.15: Techniques used in LSTM

Split data	Technique type	CV	Optimizer	Loss	Epoch	Early stopping
70% training 30% testing	Stratified sampling	10	Adam	Binary cross entropy	50	Monitor = Val _loss Mini delta = 0 Mode = auto Patience = 20

Table 5.16: Results of LSTM performance

Test case Acute asthma	Test case Acute bronchitis	Sensitivity	Specificity	ROC- AUC	Test loss	Accuracy	Time
75	79	0.9866	0.9873	0.9986	0.0487	98.7012	49 ms

Table 5.17: Number of parameters used in LSTM

Total parameters	Trainable parameters	Non trainable parameters
144,497	144,497	0

The results of test accuracy and test loss for LSTM model appeared as in Figure 5.7 and 5.8.

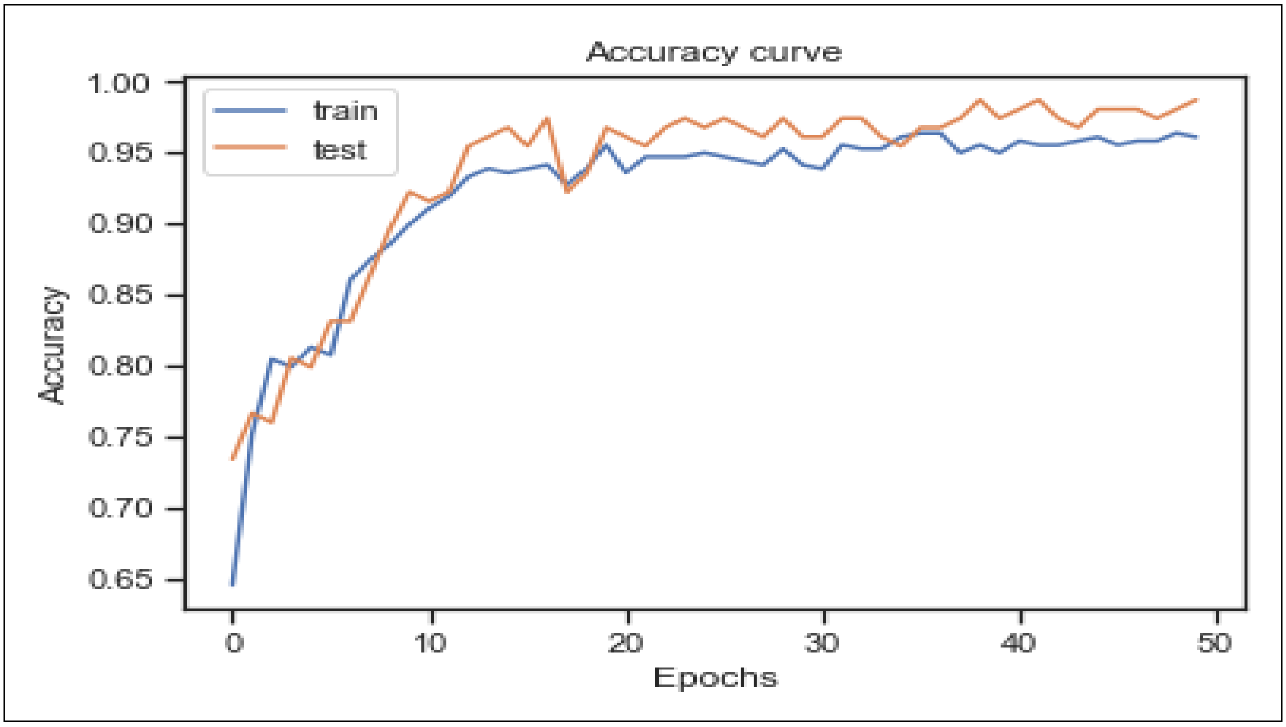

Figure 5.7: Accuracy curve of LSTM model

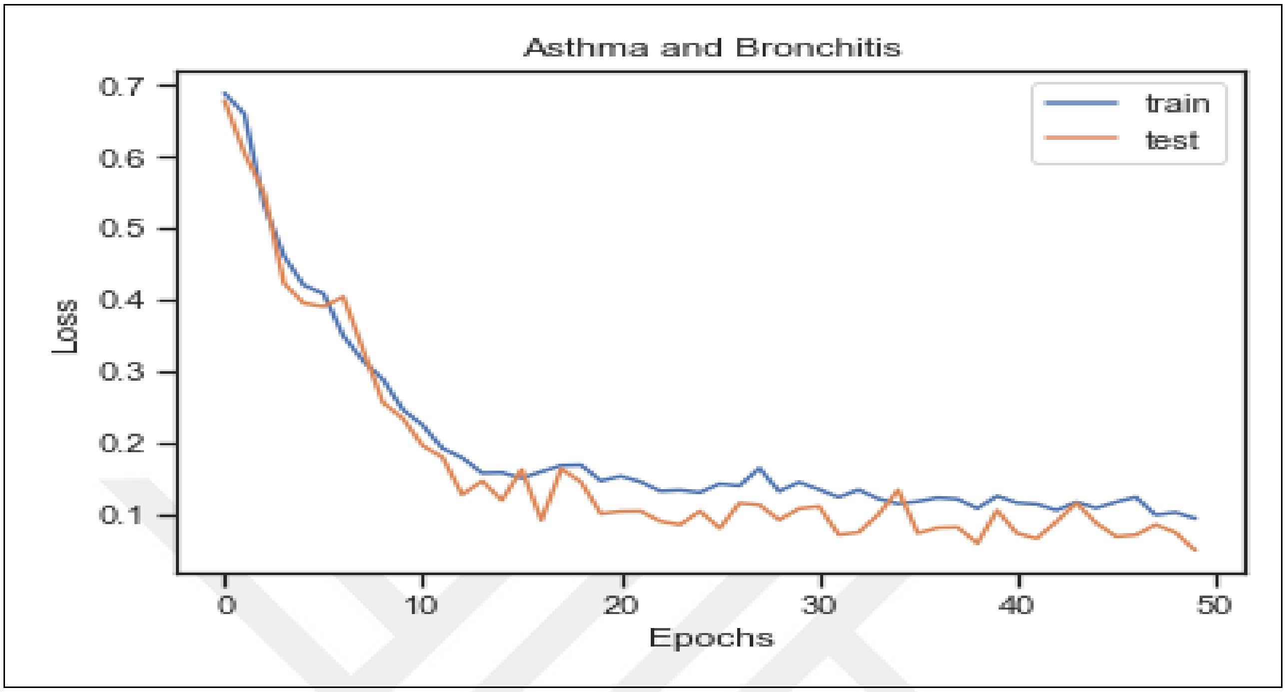

Figure 5.8: Loss curve of LSTM model

After testing the model and obtaining the final accuracy with test loss, the confusion matrix is calculated for the data set as in table 5.18 and Figure 5.9

Table 5.18: Results of LSTM confusion matrix

Class	Precision	Recall	F1-score	Support cases
Acute asthma	0.99	0.99	0.99	75
Acute bronchitis	0.99	0.99	0.99	79
Macro avg	0.99	0.99	0.99	154
Weight avg	0.99	0.99	0.99	154

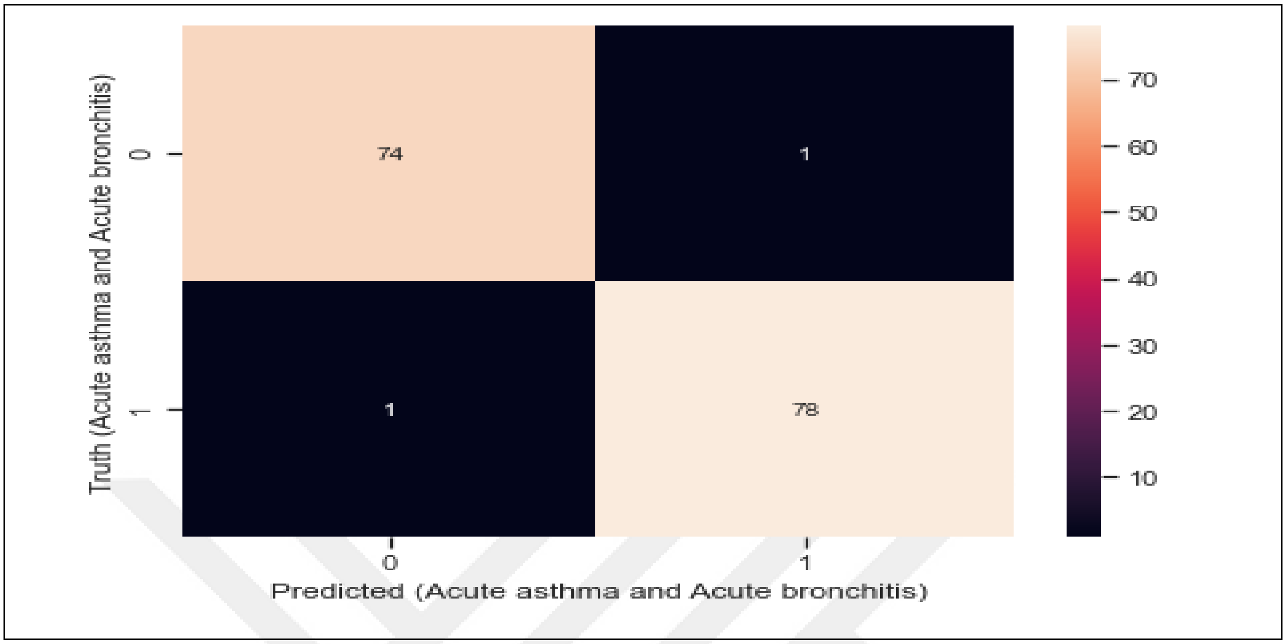

Figure 5.9: Confusion matrix of LSTM model

5.2.3 Comparing Between DL Results

After the appearance of the results of the deep learning models (1D-CNN, LSTM) and their comparison, it was found that 1D-CNN outperformed the rest of the models with an accuracy of 99.3506 to differentiate between acute asthma and bronchitis, as in table 5.19.

Table 5.19: Comparing between DL results

Classifier	ROC- AUC	Sensitivity	Specificity	F1-score	Test loss	Accuracy	Time
1D-CNN	0.9932	0.9866	1.0	0.99	0.0432	99.3506	12 ms
LSTM	0.9986	0.9873	0.9620	0.99	0.0487	98.7012	49 ms

5.3 COMPARING BETWEEN ML AND DL RESULTS

After obtaining the results for both DL and ML , we will make a final comparison of all the models used in this study and choose the model that obtained the highest accuracy, as in table 5.20.

Table 5.20: Comparing between ML and DL results

Classifier	Split data	ROC- AUC	Sensitivity	Specificity	F1-score	Accuracy
1D-CNN	70% training 30% testing	0.9932	0.9866	1.0	0.99	99.3506
MLP	70% training 30% testing	0.1	1.0	0.9746	0.99	98.7012
LSTM	70% training 30% testing	0.9986	0.9873	0.9620	0.99	98.7012
K_NN	70% training 30% testing	0.9801	1.0	0.9620	0.98	98.0519
DT	70% training 30% testing	0.9944	1.0	0.9367	0.97	96.7532

5.4 DISCUSSION OF RESULTS AND MODEL SELECTION

After training and testing both the ML and DL models, and Apply the same techniques to all models used in this study, it was found that the best model among the ML models that is MLP with accuracy of (98.7012) and ROC- AUC (0.1) as in table 5.10 , because it contains a number of hidden layers and parameters that processes of data entered in a complex mathematical way and outputting the results accurately[22]. The nearest neighbour algorithm also achieved excellent accuracy to distinguish between the two diseases, reaching (98.0519) with ROC-AUC of (0.9801), and the decision tree accuracy was (96.7532) with ROC-AUC of (0.9944) as in table 5.20 ,these results are considered desirable, and this is what was indicated during the literature review[2][19]. While Convolutional neural networks were the best model that has obtained the highest accuracy of (99.3506) and ROC-AUC (0.9932) among deep learning models as in table 5.19 because these networks are inspired by the human brain and contain a techniques that can process data a sophisticated way and show the desired results, especially in the field of health care[29] [30]. The accuracy of the LSTM was (98.7012) which is the same as that of MLP , but the ROC- AUC , sensitivity and specificity was higher than LSTM Finally, the convolutional neural network model can be approved as a binary classification model for the distinction between acute asthma and bronchitis in this study by the comparing between ML and DL models , due this model was well trained and tested as shown in the figure 5.4, also the curve loss rate was (0.0432) as in figure 5.5, in addition, the error rate in the confusion matrix was few compared with the other models as in figure 5.6 .

6. CONCLUSIONS AND FUTURE WORK

6.1 CONCLUSIONS

After completing this study and obtaining the results using ML and DL models to differentiate between acute asthma and bronchitis, we concluded the following:

a. The number of data amounting to 512 real cases is considered acceptable compared to the algorithms used in this study, due that a period of four months is not sufficient to collect larger data, in addition to that all cases concern children under school age, and each case takes a long time to examine by a paediatric consultant because 12 clinical features must be identified in order to achieve the existence of one of the two diseases .
b. The data collection period lasted from March 2022 to June 2022, when acute asthma cases increased in dusty atmospheres and air pollution such as fumes and smoke, while acute bronchitis cases increased when the temperature was low, and acute asthma amount 248 cases, while bronchitis cases were 264 cases.
c. The pre-processing of the acute asthma and bronchitis dataset was carried out in an easy and accurate way, since all of them were numerical data and after they were cleaned and analysed through the correlation matrix, and was the verified firstly with the consultant, where it was agreed to keep 12 clinical features for each case.
d. Three models of ML and two models of DL were used, where the neural networks represented by 1D-CNN and MLP were the superior in all aspects and got the highest accuracy of (99.3506) and (98.7012) with ROC -AUC (1.0), (0.9746) respectively, and the 1D-CNN were select as a binary classification model for our study.

6.2 FUTURE WORK

In future work, we recommend the following:

a. Data must be collected by specialized doctors exclusively to obtain an accurate manual diagnosis.
b. Taking the largest number of samples, whether textual or x-ray, in a period of at least eight months to be sufficient especially in winter or summer because most cases are seasonal.
c. A smart application can be made on mobile phones using convolutional neural networks that helps junior doctors to distinguish between acute asthma and bronchitis in children, especially in emergency halls.
d. In the future, work can be done on other diseases in children, such as predicting of autism, after the approval of the consultant doctor.

REFERENCES

[1] D. Patel, G. L. Hall, D. Broadhurst, A. Smith, A. Schultz, and R. E. Foong, “Does machine learning have a role in the prediction of asthma in children?,” *Paediatr. Respir. Rev.*, vol. 41, no. xxxx, pp. 51–60, 2022, doi: 10.1016/j.prrv.2021.06.002.

[2] K. Stokes *et al.*, “A machine learning model for supporting symptom-based referral and diagnosis of bronchitis and pneumonia in limited resource settings,” *Biocybern. Biomed. Eng.*, vol. 41, no. 4, pp. 1288–1302, 2021, doi: 10.1016/j.bbe.2021.09.002.

[3] Y. Tam, E. Johansson, and T. B. Mersha, “Multi-Omics Profiling Approach to Asthma : An Evolving Paradigm,” 2022.

[4] C. De Rose, S. M. Sopo, P. Valentini, R. Morello, D. Biasucci, and D. Buonsenso, “Potential Application of Lung Ultrasound in Children with Severe Uncontrolled Asthma : Preliminary Hypothesis Based on a Case Series,” 2022.

[5] J. P. David McCune, “Reducing Unnecessary Antibiotic Treatment for Acute Bronchitis Using Real-Time, Text-Based Primary Care,” *Telehealth Med. Today*, vol. 6, no. 3, pp. 1–9, 2021.

[6] Olufunke C. Olayemi * Adewale O. Sunday **, “Application of Machine Learning to The Diagnosis of Lower Respiratory Tract Infections in Paediatric Patients,” no. January, 2017.

[7] R. Kleiman *et al.*, “Machine Learning Assisted Discovery of Novel Predictive Lab Tests Using Electronic Health Record Data.,” *AMIA Jt. Summits Transl. Sci. proceedings. AMIA Jt. Summits Transl. Sci.*, vol. 2019, pp. 572–581, 2019, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/31259012%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6568080

[8] M. Kim, J. Jang, S. Jeon, and S. Youm, “applied sciences A Study on Customized Prediction of Daily Illness Risk Using Medical and Meteorological Data,” 2022.

[9] S. Shurrab and R. Duwairi, “Self-supervised learning methods and applications in medical imaging analysis: a survey,” *PeerJ Comput. Sci.*, vol. 8, 2022, doi: 10.7717/PEERJ-

CS.1045.

[10] Y. Tong, B. Lin, G. Chen, and Z. Zhang, “Predicting Continuity of Asthma Care Using a Machine Learning Model : Retrospective Cohort Study,” 2022.

[11] J. Peng, C. Chen, M. Zhou, X. Xie, Y. Zhou, and C. H. Luo, “A Machine-learning Approach to Forecast Aggravation Risk in Patients with Acute Exacerbation of Chronic Obstructive Pulmonary Disease with Clinical Indicators,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–9, 2020, doi: 10.1038/s41598-020-60042-1.

[12] Y. Raita *et al.*, “Machine learning - based prediction of acute severity in infants hospitalized for bronchiolitis : a multicenter prospective study,” *Sci. Rep.*, pp. 1–11, 2020, doi: 10.1038/s41598-020-67629-8.

[13] R. Ullah, S. Khan, H. Ali, I. I. Chaudhary, M. Bilal, and I. Ahmad, “A comparative study of machine learning classifiers for risk prediction of asthma disease,” *Photodiagnosis Photodyn. Ther.*, vol. 28, pp. 292–296, 2019, doi: 10.1016/j.pdpdt.2019.10.011.

[14] N. Shahid, T. Rappon, and W. Berta, “Applications of artificial neural networks in health care organizational decision-making: A scoping review,” *PLoS One*, vol. 14, no. 2, pp. 1–22, 2019, doi: 10.1371/journal.pone.0212356.

[15] D. Spathis and P. Vlamos, “Diagnosing asthma and chronic obstructive pulmonary disease with machine learning,” *Health Informatics J.*, vol. 25, no. 3, pp. 811–827, 2019, doi: 10.1177/1460458217723169.

[16] J. L. Harvey and S. A. P. Kumar, “Machine Learning for Predicting Development of Asthma in Children Machine Learning for Predicting Development of Asthma in Children,” no. December 2019, 2020, doi: 10.1109/SSCI44817.2019.9002692.

[17] A. P. Pawlovsky, “An ensemble based on distances for a kNN method for heart disease diagnosis,” *Int. Conf. Electron. Inf. Commun. ICEIC 2018*, vol. 2018-Janua, pp. 1–4, 2018, doi: 10.23919/ELINFOCOM.2018.8330570.

[18] Y. C. Liu *et al.*, “Evaluation of the Need for Intensive Care in Children with Pneumonia:

Machine Learning Approach," *JMIR Med. Informatics*, vol. 10, no. 1, pp. 1–11, 2022, doi: 10.2196/28934.

[19] T. Cao, Y. Zhu, and Y. Zhu, "Construction of Prediction Model of Renal Damage in Children with Henoch-Schönlein Purpura Based on Machine Learning," vol. 2022, 2022.

[20] Y. Feng, Y. Wang, C. Zeng, and H. Mao, "Artificial intelligence and machine learning in chronic airway diseases: Focus on asthma and chronic obstructive pulmonary disease," *Int. J. Med. Sci.*, vol. 18, no. 13, pp. 2871–2889, 2021, doi: 10.7150/ijms.58191.

[21] S. Mishra, H. K. Tripathy, and P. K. Mallick, "EAGA-MLP — An Enhanced and Adaptive Hybrid," pp. 1–31, 2020.

[22] S. Jahan *et al.*, "Automated invasive cervical cancer disease detection at early stage through suitable machine learning model," *SN Appl. Sci.*, vol. 3, no. 10, 2021, doi: 10.1007/s42452-021-04786-z.

[23] A. Al Bataineh and S. Manacek, "MLP-PSO Hybrid Algorithm for Heart Disease Prediction," *J. Pers. Med.*, vol. 12, no. 8, p. 1208, 2022, doi: 10.3390/jpm12081208.

[24] 4 Gang Yu1 , Zheming Li1 , Shuxian Li2 , Jingling Liu2 , Moyuan Sun3 , Xiaoqing Liu3 , Fenglei Sun3 , Jie Zheng3 , Yiming Li3 , Yizhou Yu3 , Qiang Shu4 , Yingshuo Wang2, "The role of artificial intelligence in identifying asthma in pediatric inpatient setting," vol. 8, no. 21, 2020, doi: 10.21037/atm-20-2501a.

[25] L. F. Samhan, A. H. Alfarra, S. S. Abu-Naser, and I. A. Amassi, "Classification of Alzheimer's Disease Using Convolutional Neural Networks," *Int. J. Acad. Inf. Syst. Res.*, vol. 6, no. 3, pp. 18–23, 2022, [Online]. Available: www.ijeais.org/ijaisr

[26] V. R. Allugunti, "A machine learning model for skin disease classification using convolution neural network," *Int. J. Comput. Program. Database Manag.*, vol. 3, no. 1, pp. 141–147, 2022.

[27] A. Srivastava, S. Jain, R. Miranda, S. Patil, S. Pandya, and K. Kotecha, "Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease,"

PeerJ Comput. Sci., vol. 7, pp. 1–22, 2021, doi: 10.7717/PEERJ-CS.369.

[28] C. Bales *et al.*, "Can machine learning be used to recognize and diagnose coughs?," *2020 8th E-Health Bioeng. Conf. EHB 2020*, pp. 19–22, 2020, doi: 10.1109/EHB50910.2020.9280115.

[29] C. Chen, Z. Hua, R. Zhang, G. Liu, and W. Wen, "Automated arrhythmia classification based on a combination network of CNN and LSTM," *Biomed. Signal Process. Control*, vol. 57, p. 101819, 2020, doi: 10.1016/j.bspc.2019.101819.

[30] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, p. 107398, 2021, doi: 10.1016/j.ymssp.2020.107398.

[31] P. N. Srinivasu, J. G. Sivasai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, "Networks with MobileNet V2 and LSTM," pp. 1–27, 2021.

[32] T. D. Pham, "Time–frequency time–space LSTM for robust classification of physiological signals," *Sci. Rep.*, vol. 11, no. 1, pp. 1–11, 2021, doi: 10.1038/s41598-021-86432-7.

[33] F. Demir and B. Taşcı, "An Effective and Robust Approach Based on R-CNN+LSTM Model and NCAR Feature Selection for Ophthalmological Disease Detection from Fundus Images," *J. Pers. Med.*, vol. 11, no. 12, 2021, doi: 10.3390/jpm11121276.

[34] E. Balaji, D. Brindha, V. K. Elumalai, and R. Vikrama, "Automatic and non-invasive Parkinson's disease diagnosis and severity rating using LSTM network," *Appl. Soft Comput.*, vol. 108, p. 107463, 2021, doi: 10.1016/j.asoc.2021.107463.

[35] M. G. El-shafiey, A. Hagag, and M. A. Ismail, "A Hybrid Bidirectional LSTM and 1D CNN for Heart Disease Prediction," vol. 21, no. 10, 2021, doi: 10.22937/IJCSNS.2021.21.10.18.

[36] M. Djerioui, Y. Brik, M. Ladjal, and B. Attallah, "Heart Disease prediction using MLP and LSTM models," *2020 Int. Conf. Electr. Eng. ICEE 2020*, 2020, doi: 10.1109/ICEE49691.2020.9249935.

[37] S. Muthukumarasamy, A. K. Tamilarasan, J. Ayeelyan, and M. Adimoolam, “Machine learning in healthcare diagnosis,” *Blockchain Mach. Learn. e-Healthcare Syst.*, pp. 343–366, 2021, doi: 10.1049/pbhe029e_ch14.

[38] N. Burkart and M. F. Huber, “A survey on the explainability of supervised machine learning,” *J. Artif. Intell. Res.*, vol. 70, pp. 245–317, 2021, doi: 10.1613/JAIR.1.12228.

[39] A. Alanazi, “Using machine learning for healthcare challenges and opportunities,” *Informatics Med. Unlocked*, vol. 30, no. February, p. 100924, 2022, doi: 10.1016/j.imu.2022.100924.

[40] S. M. Kumar and T. Kumanan, “Skin Lesion Classification System Using Shearlets,” *Comput. Syst. Sci. Eng.*, vol. 44, no. 1, pp. 833–844, 2022, doi: 10.32604/csse.2023.022385.

[41] M. M. Al Anazi and O. R. Shahin, “A Machine Learning Model for the Identification of the Holy Quran Reciter Utilizing K-Nearest Neighbor and Artificial Neural Networks,” *Inf. Sci. Lett.*, vol. 11, no. 4, pp. 1093–1102, 2022, doi: 10.18576/isl/110410.

[42] S. Ray, “A Quick Review of Machine Learning Algorithms,” *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019*, pp. 35–39, 2019, doi: 10.1109/COMITCon.2019.8862451.

[43] B. Charbuty and A. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning,” *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.

[44] L. Tian, Z. Zeng, and L. Ge, “Design and implementation of a decision tree classification algorithm for radar signal recognition,” no. July, p. 57, 2022, doi: 10.1117/12.2642621.

[45] Z. Sabouri, Y. Maleh, and N. Gherabi, “Benchmarking Classification Algorithms for Measuring the Performance on Maintainable Applications,” *Lect. Notes Networks Syst.*, vol. 357 LNNS, no. January, pp. 173–179, 2022, doi: 10.1007/978-3-030-91738-8_17.

[46] Y. chen Wu and J. wen Feng, “Development and Application of Artificial Neural Network,” *Wirel. Pers. Commun.*, vol. 102, no. 2, pp. 1645–1656, 2018, doi: 10.1007/s11277-017-

5224-x.

[47] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. E. Mohamed, and H. Arshad, “State-of-the-art in artificial neural network applications: A survey,” *Heliyon*, vol. 4, no. 11, p. e00938, 2018, doi: 10.1016/j.heliyon.2018.e00938.

[48] ChigNwankpaozie, W. Ijomah, A. Gachagan, and S. Marshall, “Activation Functions: Comparison of trends in Practice and Research for Deep Learning,” pp. 1–20, 2018, [Online]. Available: http://arxiv.org/abs/1811.03378

[49] R. Jin and Q. Niu, “Automatic Fabric Defect Detection Based on an Improved YOLOv5,” *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/7321394.

[50] S. Narang and S. Kumar, “" Neural Network : Basics and Application ”,” vol. 11, no. 2, pp. 95–99, 2022.

[51] E. Pashaei and E. Pashaei, “Training Feedforward Neural Network Using Enhanced Black Hole Algorithm: A Case Study on COVID-19 Related ACE2 Gene Expression Classification,” *Arab. J. Sci. Eng.*, vol. 46, no. 4, pp. 3807–3828, 2021, doi: 10.1007/s13369-020-05217-8.

[52] L. Alzubaidi *et al.*, *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00444-8.

[53] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, “Review on Convolutional Neural Networks (CNN) in vegetation remote sensing,” *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 24–49, 2021, doi: 10.1016/j.isprsjprs.2020.12.010.

[54] A. Patil and M. Rane, “Convolutional Neural Networks: An Overview and Its Applications in Pattern Recognition,” *Smart Innov. Syst. Technol.*, vol. 195, pp. 21–30, 2021, doi: 10.1007/978-981-15-7078-0_3.

[55] Z. Li, W. Yang, S. Peng, and F. Liu, “CNN survey,” 2020.

[56] Y. Wang *et al.*, "Unsupervised machine learning for the discovery of latent disease clusters and patient subgroups using electronic health records," *J. Biomed. Inform.*, vol. 102, no. December 2019, p. 103364, 2020, doi: 10.1016/j.jbi.2019.103364.

[57] Y. Okada *et al.*, "Clinical Phenotyping of Out-of-Hospital Cardiac Arrest Patients with Shockable Rhythm — Machine Learning-Based Unsupervised Cluster Analysis —," *Circ. J.*, vol. 86, no. 4, pp. 668–676, 2022, doi: 10.1253/circj.CJ-21-0675.

[58] A. Chebli, A. Djebbar, and H. F. Marouani, "Semi-Supervised Learning for Medical Application: A Survey," *Proc. 2018 Int. Conf. Appl. Smart Syst. ICASS 2018*, no. November, pp. 1–9, 2019, doi: 10.1109/ICASS.2018.8651980.

[59] K. Wenger *et al.*, "A semi-supervised learning approach for bladder cancer grading," *Mach. Learn. with Appl.*, vol. 9, no. June, p. 100347, 2022, doi: 10.1016/j.mlwa.2022.100347.

[60] MarcelPanzer and B. Bender, "Deep reinforcement learning in production systems: a systematic literature review," *Int. J. Prod. Res.*, vol. 60, no. 13, pp. 4316–4341, 2022, doi: 10.1080/00207543.2021.1973138.

[61] P. P. Shinde and S. Shah, "A Review of Machine Learning and Deep Learning Applications," *Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018*, pp. 1–6, 2018, doi: 10.1109/ICCUBEA.2018.8697857.

[62] H. van der Heijden, "Predicting industry sectors from financial statements: An illustration of machine learning in accounting research," *Br. Account. Rev.*, no. April 2020, 2022, doi: 10.1016/j.bar.2022.101096.

[63] P. Mathur, "Machine learning applications using python: Cases studies from healthcare, retail, and finance," *Mach. Learn. Appl. Using Python Cases Stud. from Heal. Retail. Financ.*, pp. 1–379, 2018, doi: 10.1007/978-1-4842-3787-8.

[64] S. Mukherjee, B. Ghosh, and K. Sitaraman, "Understanding patterns and structures in data--Analytics and Machine Learning and application to Travel Industry," *Researchgate.Net*, no. November 2013, 2014.

[65] Y. Akgül and A. O. Uymaz, *Facebook/Meta usage in higher education: A deep learning-based dual-stage SEM-ANN analysis*. Springer US, 2022. doi: 10.1007/s10639-022-11012-9.

[66] K. Ghennam, F. Attou, and F. Abdoun, "Impact of atmospheric pollution on asthma and bronchitis based on lichen biomonitoring using IAP, IHI and GIS in Algiers Bay (Algeria)," *Environ. Monit. Assess.*, vol. 193, no. 4, pp. 1–12, 2021, doi: 10.1007/s10661-021-08965-8.

[67] A. Ijaz *et al.*, "Informatics in Medicine Unlocked Towards using cough for respiratory disease diagnosis by leveraging Artificial Intelligence : A survey," *Informatics Med. Unlocked*, vol. 29, p. 100832, 2022, doi: 10.1016/j.imu.2021.100832.

[68] Ö. Çapraz and A. Deniz, "Assessment of hospitalizations from asthma, chronic obstructive pulmonary disease and acute bronchitis in relation to air pollution in İstanbul, Turkey," *Sustain. Cities Soc.*, vol. 72, no. May, 2021, doi: 10.1016/j.scs.2021.103040.

[69] D. Mitra, A. Paul, and S. Chatterjee, "Machine learning in healthcare," *AI Innov. Med. Imaging Diagnostics*, pp. 37–60, 2021, doi: 10.4018/978-1-7998-3092-4.ch002.

[70] P. Misra and A. S. Yadav, "Impact of Preprocessing Methods on Healthcare Predictions," *SSRN Electron. J.*, no. July, 2019, doi: 10.2139/ssrn.3349586.

[71] A. Y. Taha, S. Tiun, A. H. A. Rahman, and A. Sabah, "Multilabel Over-sampling and Under-sampling with Class Alignment for Imbalanced Multilabel Text Classification," *J. Inf. Commun. Technol.*, vol. 20, no. 3, pp. 423–456, 2021, doi: 10.32890/JICT2021.20.3.6.

[72] F. Wu, X. Liu, Y. Wang, X. Li, and M. Zhou, "Research on Evaluation Model of Hospital Informatization Level Based on Decision Tree Algorithm," *Secur. Commun. Networks*, vol. 2022, 2022, doi: 10.1155/2022/3777474.

[73] V. Chaurasia and S. Pal, "Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer," *SN Comput. Sci.*, vol. 1, no. 5, 2020, doi: 10.1007/s42979-020-00296-8.

[74] E. van den Heuvel and Z. Zhan, "Myths About Linear and Monotonic Associations: Pearson's r, Spearman's ρ, and Kendall's τ," *Am. Stat.*, vol. 76, no. 1, pp. 44–52, 2022, doi: 10.1080/00031305.2021.2004922.

[75] A. Jesu, P. Lazar, D. K. Sharma, and T. Guna, "Fetal Health Classification from Cardiotocographic Data Using Machine Learning," no. March 2022, 2021.

[76] S. BUYRUKOĞLU and A. AKBAŞ, "Machine Learning based Early Prediction of Type 2 Diabetes: A New Hybrid Feature Selection Approach using Correlation Matrix with Heatmap and SFS," *Balk. J. Electr. Comput. Eng.*, vol. 10, no. 2, pp. 110–117, 2022, doi: 10.17694/bajece.973129.

[77] D. S. Soper, "Greed is good: Rapid hyperparameter optimization and model selection using greedy k-fold cross validation," *Electron.*, vol. 10, no. 16, 2021, doi: 10.3390/electronics10161973.

[78] B. G. Marcot and A. M. Hanea, "What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?," *Comput. Stat.*, vol. 36, no. 3, pp. 2009–2031, 2021, doi: 10.1007/s00180-020-00999-9.

[79] S. Prusty, S. Patnaik, and S. K. Dash, "SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer," *Front. Nanotechnol.*, vol. 4, no. August, pp. 1–12, 2022, doi: 10.3389/fnano.2022.972421.

[80] T. A. Assegie, "An optimized K-Nearest neighbor based breast cancer detection," *J. Robot. Control*, vol. 2, no. 3, pp. 115–118, 2021, doi: 10.18196/jrc.2363.

[81] S. Lee, C. Lee, K. G. Mun, and D. Kim, "Decision Tree Algorithm Considering Distances between Classes," *IEEE Access*, vol. 10, no. June, pp. 69750–69756, 2022, doi: 10.1109/ACCESS.2022.3187172.

[82] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, pp. 1–13, 2020, doi: 10.1186/s12864-019-6413-7.

[83] P. Wang, L. Song, Y. Hao, H. Wang, S. Li, and L. Cui, “A light intelligent diagnosis model based on improved Online Dictionary Learning sample-making and simplified convolutional neural network,” *Measurement*, vol. 183, no. 52075030, p. 109813, 2021, doi: 10.1016/j.measurement.2021.109813.

APPENDIX A

ALL THE CODES USED IN PYTHON – JUPYTER NOTEBOOK

A.1 MACHINE LEARNING CODE

```
# Import libraries
import numpy as np
import pandas as pd
import matplotlib. pyplot as plt
import seaborn as sns; sns.set(style="ticks", color_codes=True)
*****************************************************************
# Read of dataset
dataset = pd. read_csv('data.csv')
dataset.head(5)
dataset['Clinical Finding'].unique()
dataset['Clinical Finding'].value_counts(normalize=True)*100
dataset.describe()
*****************************************************************

# Split our dataset into 70% training  and  30% testing
from sklearn.model_selection import train_test_split
x = dataset.drop('Clinical Finding', axis=1)
y = dataset['Clinical Finding']
Xtrain,  Xtest,  Ytrain,  Ytest  =  train_test_split(x,  y,  test_size
=0.3,random_state=0)
 *****************************************************************

# Standardization of dataset
from sklearn.preprocessing import StandardScaler
sc=StandardScaler()
scaler = sc.fit(Xtrain)
trainX_scaled = scaler.transform(Xtrain)
testX_scaled = scaler.transform(Xtest)
*****************************************************************

# Coloration Matrix of dataset
import matplotlib.pyplot as plt
import seaborn as sns; sns.set(style="ticks", color_codes=True)
plt.figure(figsize= [12,15])
sns.heatmap(dataset.corr(), annot=True, cmap="YlGnBu")
plt.title('Correlations of the Features')
plt.show()
*****************************************************************
```

```
# Build Classifier of K - Nearest Neighbours (K-NN)
from sklearn.model_selection import GridSearchCV
k_range = list(range(1, 31))
param_grid = dict(n_neighbors=k_range)
*****************************************************************

# Defining parameter range
grid=GridSearchCV(knn,param_grid,cv=10,scoring='accuracy',
return_train_score=False,verbose=1)
*****************************************************************

# Fitting model for grid search
grid_search=grid.fit(Xtrain, Ytrain)
print(grid_search.best_params_)
print(k_range)
*****************************************************************

# Build Classifier of Decision Tree (DT)
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import GridSearchCV
Hyper_paramters={'criterion':['gini',
'entropy'],'max_depth':[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,1
9,20]}
Tree_Gridsearch_paramters=GridSearchCV(DecisionTreeClassifier(),Hyper_
paramters, scoring='roc_auc',n_jobs=-1,cv=10,verbose=1)
Tree_crossvalidation=Tree_Gridsearch_paramters.fit(X_train,y_train)
print ("The best paramter combination is ")
print(Tree_crossvalidation.best_params_)
Final_Model=Tree_crossvalidation.best_estimator_
print("The best AUC score was ")
print(Tree_crossvalidation.best_score_)

*****************************************************************

# Build Classifier of Multi-Layer Perceptron (MLP)
from sklearn.neural_network import MLPClassifier
from sklearn.model_selection import GridSearchCV
Hyper_paramters={'batch_size':["auto",100],'hidden_layer_sizes':[(5,5,
5),(5),(6),(4)],'learning_rate_init':[0.001,0.005,
0.01],'max_iter':[500,600]}
NN_Gridsearch_paramters=GridSearchCV(MLPClassifier(),Hyper_paramters,s
coring='roc_auc',n_jobs=-1,cv=10,return_train_score=False,verbose=0)
NN_Gridsearch_paramters=GridSearchCV(MLPClassifier(early_stopping=Fals
e,random_state=0),Hyper_paramters,scoring='roc_auc',n_jobs=-
1,cv=10,verbose=0)
```

```
NN_crossvalidation=NN_Gridsearch_paramters.fit(trainX,trainY)
print ("The best paramter combination is ")
print(NN_crossvalidation.best_params_)  #gets the best estimator
Final_Model=NN_crossvalidation.best_estimator_
print("The best AUC score was ")
print(NN_crossvalidation.best_score_)
```
****

```
# Confusion Matrix of ML
y_pred = Final_Model.predict(X_test)
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
cm1 = confusion_matrix(y_test, y_pred)
print(cm1)
print(classification_report(y_test,   y_pred,   target_names=["   Acute
asthma", " Acute bronchitis"]))

%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sn
plt.figure(figsize=(7,5))
sn.heatmap(cm1, annot=True)
plt.xlabel('Predicted (Acute asthma and Acute bronchitis)')
plt.ylabel('Truth (Acute asthma and Acute bronchitis)')
```
****

```
# Calculate the sensitivity and specificity of ML
total1=sum(sum(cm1))
accuracy1=(cm1[0,0]+cm1[1,1])/total1
print ('Accuracy : ', accuracy1)
sensitivity1 = cm1[0,0]/(cm1[0,0]+cm1[0,1])
print('Sensitivity : ', sensitivity1 )
specificity1 = cm1[1,1]/(cm1[1,0]+cm1[1,1])
print('Specificity : ', specificity1)
```

****

A.1.1 Deep Learning Code

```
# Build Classifier of Convolutional Neural Network (CNN)
# Import Libraries
import tensorflow as tf
from tensorflow import keras
from keras import Sequential
from keras.layers import Flatten,Dense,Conv1D,MaxPool1D,Dropout
from keras.optimizers import Adam
import numpy as np
import pandas as pd
*****************************************************************

# Read of dataset
dataset = pd. read_csv('data.csv')
dataset.head(5)
dataset['Clinical Finding'].unique()
dataset['Clinical Finding'].value_counts(normalize=True)*100
dataset.describe()
*****************************************************************

# Split our dataset into 70% training and 30% testing
from sklearn.model_selection import train_test_split
x = dataset.drop('Clinical Finding', axis=1)
y = dataset['Clinical Finding']
Xtrain, Xtest, Ytrain, Ytest = train_test_split(x, y, test_size =
0.3,random_state=0 , stratify=y )
*****************************************************************

# Standardization of dataset
from sklearn.preprocessing import StandardScaler
sc=StandardScaler()
scaler = sc.fit(Xtrain)
trainX_scaled = scaler.transform(Xtrain)
testX_scaled = scaler.transform(Xtest)
*****************************************************************

# Build model of CNN
model = Sequential()
model.add(Conv1D(32,kernel_size=3,
activation='relu',input_shape=(12,1)))
model.add(Dropout(0.1))
model.add(Conv1D(64,3 , activation='relu'))
model.add(Dropout(0.2))
model.add(Conv1D(128, 3, activation='relu'))
```

```
model.add(Dropout(0.5))
model.add(MaxPool1D(pool_size=2))

model.add(Flatten())

model.add(Dense(128, activation='relu'))
model.add(Dropout(0.3))

model.add(Dense(1,activation='sigmoid'))

model.compile(optimizer='Adam',loss='binary_crossentropy',metrics=['ac
curacy','AUC'])
*****************************************************************
```

```
# Build Classifier of Long Short-Term Memory (LSTM)
# Import Libraries
import tensorflow as tf
import numpy as np
import pandas as pd
from   tensorflow.keras.layers import LSTM
from   tensorflow.keras.models import Sequential
from   tensorflow.keras.layers import Dense,Dropout
*****************************************************************
```

```
# Build model OF LSTM
model = Sequential()
model.add(LSTM(100, input_shape=(12,1),return_sequences=True))
model.add(Dropout(0.2))
model.add(LSTM(100,activation='relu'))
model.add(Dropout(0.2))
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.25))
model.add(Dense(64, activation='relu'))
model.add(Dropout(0.2))
model.add(Dense(32, activation='relu'))
model.add(Dropout(0.1))
model.add(Dense(1, activation='sigmoid'))
model.compile(optimizer='Adam',loss='binary_crossentropy',metrics=['ac
curacy','AUC'])
print(model.summary())
*****************************************************************
```

```
# Print results of CNN and LSTM
from keras import callbacks
from keras.callbacks import EarlyStopping
from keras.callbacks import ModelCheckpoint
history=model.fit(X_train,Y_train,batch_size=64,epochs=50,verbose=1,va
lidation_data=(X_test,Y_test))
checkpoint=ModelCheckpoint(filepath='best_model.h5',
monitor='val_loss',
save_best_only=True,save_weight_only=False,verbose=1,mode='auto')
early=EarlyStopping(monitor='val_loss',min_delta=0,mode='auto',verbose
=1, patience=20)
# Draw test curve and loss curve (CNN, LSTM)
plt.plot(history.history["accuracy"])
plt.plot(history.history["val_accuracy"])
plt.xlabel("Epochs")
plt.ylabel("Accuracy")
plt.title("Accuracy curve")
plt.legend(['train','test'])
plt.show()

plt.plot(history.history["loss"])
plt.plot(history.history["val_loss"])
plt.xlabel("Epochs")
plt.ylabel("Loss")
plt.title("Loss curve")
plt.legend(['train','test'])
plt.show()

*****************************************************************

# Confusion Matrix of DL
y_pred = Final_Model.predict(X_test)
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
cm1 = confusion_matrix(y_test, y_pred)
print(cm1)
print(classification_report(y_test,   y_pred,   target_names=["   Acute
asthma", " Acute bronchitis"]))

%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sn
plt.figure(figsize=(7,5))
sn.heatmap(cm1, annot=True)
plt.xlabel('Predicted (Acute asthma and Acute bronchitis)')
plt.ylabel('Truth (Acute asthma and Acute bronchitis)')
```

```
# Calculate the sensitivity and specificity of DL
total1=sum(sum(cm1))
accuracy1=(cm1[0,0]+cm1[1,1])/total1
print ('Accuracy : ', accuracy1)

sensitivity1 = cm1[0,0]/(cm1[0,0]+cm1[0,1])
print('Sensitivity : ', sensitivity1 )

specificity1 = cm1[1,1]/(cm1[1,0]+cm1[1,1])
print('Specificity : ', specificity1)
```

****