A NOVEL FRAMEWORK FOR AIRBORNE DISEASES MONITORING AND PREDICTION USING MACHINE LEARNING

Thesis Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY

in

Computer Science And Engineering

By

Sapna Kumari

Registration Number: 41800656

Supervised By

Dr. Harpreet Kaur (27633)

Professor

School of Computer Science And Engineering

Lovely Professional University, Punjab

LOVELY PROFESSIONAL UNIVERSITY, PUNJAB

2025

Declaration

I, hereby declared that the presented work in the thesis entitled “A Novel Framework for Airborne Diseases Monitoring and Prediction using Machine Learning” in fulfilment of degree of **Doctor of Philosophy (Ph.D.)** is outcome of research work carried out by me under the supervision of Dr. Harpreet Kaur, working as Professor, in the School of Computer Science and Engineering of Lovely Professional University, Punjab, India. In keeping with general practice of reporting scientific observations, due acknowledgements have been made whenever work described here has been based on findings of other investigator. This work has not been submitted in part or full to any other University or Institute for the award of any degree.

Sapna

Sapna Kumari
Registration No. : 41800656
School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab, India

Certificate

This is to certify that the work reported in the Ph.D. thesis entitled "A Novel Framework for Airborne Diseases Monitoring and Prediction using Machine Learning" submitted in fulfillment of the requirement for the reward of degree of **Doctor of Philosophy (Ph.D.)** in the School of Computer Science and Engineering, is a research work carried out by Sapna Kumari, Registration No. 41800656, is bonafide record of her original work carried out under my supervision and that no part of thesis has been submitted for any other degree, diploma or equivalent course.

Harpreet

Dr. Harpreet Kaur, Professor
UID: 27633
School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab, India

Abstract

Nowadays, the healthcare system plays a significant role in our daily life. As per the growing population, it is a big challenge for the healthcare system to manage and prevent the spread of Airborne diseases. Over the past few years, airborne illnesses have significantly increased morbidity and disability rates globally. It is a very tremendous task to monitor the patient regular basis. In the last few years, AB coronavirus diseases spread at an alarming rate.

Airborne Diseases are a worldwide concern economically and socially. Such types of diseases have sparked a big threat globally. These diseases are spread when an infected person sneezing, talking, coughing, laughing, or singing. ABDs can be spread directly or indirectly. Human contact and air droplets are the primary means of communication for the spreading of airborne illnesses. Human-to-human transmission can result from direct contact or indirect contact via an intermediate object or by touching the surface. Tuberculosis, Coronavirus, Influenza, Measles, Encephalitis, Diphtheria, and Chickenpox are commonly major airborne diseases. Tuberculosis and Measles virus (rubella) require a long distance to transmit while Whooping cough, Chickenpox, and Mumps virus are additional ABDs which can transmit over a short route distance.

Airborne diseases are a vast area of research in which a large number of diseases are covered. But the focus of this research is limited to coronavirus disease and its variant Omicron as it has been one of the most prevailing during recent years. This study aims to propose an efficient prediction model for diagnosing, monitoring, and identifying the coronavirus and its variant omicron cases at an early stage. In 2020 coronavirus had spread in many countries that affected severely to the person health, caused millions of deaths in results.

In the current time, it is a challenging task for the healthcare industry and government agencies to accurately detect the disease at the initial stage. In the current era of development, ML which is part of AI provides enormous services in the field of healthcare. Obsolete healthcare technology is not able to monitor and prevent the spread of such diseases efficiently. Therefore, in the current research work, an efficient framework is proposed for monitoring and preventing the spread of airborne disease using machine learning approaches over spatio-temporal manner. In this thesis, the deep learning part of ML, a SQueezeNet framework has been used for predicting the coronavirus disease. The proposed approach has produced 97.4% accuracy and for four-class and 98.4% accuracy for three-class classification respectively. The value of other statistical parameters in terms of specificity is 91.0% and the f-measure consists of 95.1% which is much better than other approaches. Moreover, for timely decision-making and the early detection and monitoring the coronavirus patients. The current study uses the enormous potential of IoT-based technology for identifying omicron. This research has also been presented a prediction model using fog-cloud computing for monitoring, diagnosing, and predicting the omicron on time. The weighted naive bayes algorithm has been used for the classification and analysis of the health state vulnerability of an individual. In addition, in the proposed model Long Short-Term-Memory and Enhanced temporal data-based Re-current neural network (LSTM-ERNN) are utilized for combating and preventing the spread of such disease. The various experimental results have shown the proposed system obtained a higher accuracy of 98.7%, recall consists of 97.8%, a precision value 92.8%, and the f-measure consists of 97.4%. Furthermore, the evaluation of the effectiveness of the prediction model has been performed in terms of Root Mean Square error (RSE) and Squared Pearson correlation coefficient (SPCRC) which has achieved less error as compared with other prediction models. The experimental results concluded that the presented system has produced better results when compared to other state-of-the-art prediction models.

Acknowledgements

From the core of my heart, I would like to express my deepest gratitude to God. He has given me strength and encouragement through all the rough and tough times during this research journey.

I would like to express my gratitude to my supervisor, **Dr. Harpreet Kaur**, for her supervision, guidance, advice, and encouragement which helped me tremendously in the completion of this thesis. Because of her guidance, encouragement, and persistent support, I would have been to complete my thesis. She always helped me and gave me extraordinary experiences that have been a substantial aid in this accomplishment.

I am grateful to all my friends and fellow researchers, especially Dr. Munish Bhatia and Dr. Prabal Gupta for their insightful suggestions and words that helped me in the completion of my work in a significant way.

I would like to express my thanks to my parents, especially my father who encouraged and supported me every time. I am thankful to my family members for their tremendous support, encouragement, and prayer for showing concern. Their constant prayers and blessings for me and my progress during this entire journey have allowed me to successfully complete my thesis.

I am grateful to extend my gratitude, to the Hon'ble Principal, BBK DAV College for Women, Amritsar for their support and for letting me be involved in the research study.

I would like to grab this opportunity to express my heartful thanks to my husband, Dinesh Gupta, for his unwavering support and encouragement throughout this journey. My life, my loving Son, **Vivaan Gupta**, deserves particular thanks for his sacrifice throughout this journey. This has been an overwhelming experience, and I am unsure how I can adequately express my gratitude to my family members.

Last but not least, I thank God for sailing me through all the rough and tough times during this research work.

Sapna Kumari

Contents

List of Figures

List of Tables

Abbreviations

WHO	World Health Organization
ABD	Airborne Disease
SARS	Severe Acute Rspiratory Syndrome
MERS	Middle East Respiratory Syndrome
TB	Tuberculosis
CorV	Coronavirus
SARS-CoV-2	Severe Acute Rspiratory Syndrome Coronavirus 2
CT	Computed Tomography
MRI	Magnetic Resonance Imaging
ML	Machine Learning
SVM	Support Vector Machine
PCA	Principal Component Analysis
SC	Stochastic Gradient
DQN	Deep Q-networks
ARIMA	Auto-regressive Integrated Moving Average
LSTM	Long Short Term Memory
ST	Spatio-temporal
AES	Acute Encephalitis Syndrome
RNN	Recurrent Neural Network
CNN	Convolutional Neural Network
MXP	Max Pooling
GAP	Global Average Pooling

PL	Pooling Layer
ReLU	Rectified Linear Unit
FCL	Fully Connected and Activation Function Layer
SQNet	SqueezeNet
TL	Transfer Learning
SMOTE	Synthetic Minority Over-sampling Technique
PCA	Principal Component Analysis
Matlab	Matrix of Laboratory
VGG	Visual Geometry Group
SSL	Secure Socket Layer
TSL	Transport security Layer
ACR	Accuracy
TR	Truthfulness
SPFC	Specificity
FLT	Faultless
NPV	Negative Predicted Value
MCR	Matthew Correlation Coefficient
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
PPV	Positive Predicted Value
ADAM	Adaptive Moment Estimation
SPT	Spatio-temporal
MNK	Minimum Number of Occurrences
TNK	Total Number of Keywords
RFR	Random Forest Regression
SK	Selected keywords
DSAPP	Data Sensation and Pre-processing Phase

DCP	Data Classification Phase
SOA	State-of-the-art
DFP	Data Filtering Phase
DPP	Data Prediction Phase
VADMP	Visualization and Decision-Making Phase
SVR	Support Vector Regression
GPS	Global positioning systems
NPV	Negative predictive value
ANN	Artificial Neural Network
SVR	Support Vector Regression
RFR	Random Forest Regression
RSE	Root Mean Square error
SPCRC	Squared Pearson correlation coefficient

Dedicated to my beloved family…

Chapter 1

Introduction

1.1 Overview Of Airborne Diseases

Nowadays, healthcare industry plays a significant role in our daily life. As per the rise in population and social interactions, healthcare systems face challenges for managing the healthcare system. Airborne diseases (ABD) is constitute a major threat all over the world, leads to sometimes death [1]. The prevalence of airborne diseases increases the chronically ill people across the world. Influenza, Tuberculosis, Measles, Encephalitis Coronavirus, and Mumps are deadly airborne diseases which can be caused by bacteria, fungi or viruses [2]. Such diseases are rapidly spread by tiny pathogens and transmits through:

(i) an infected person to another person;

(ii) or via through an environment events.

Many studies have shown that Airborne Disease diseases like Tuberculosis (TB), SARS-CoV-2 (COVID-19), and Avian Influenza are serious illnesses that spread quickly and have a negative impact on an individual's health [3]. Human airborne

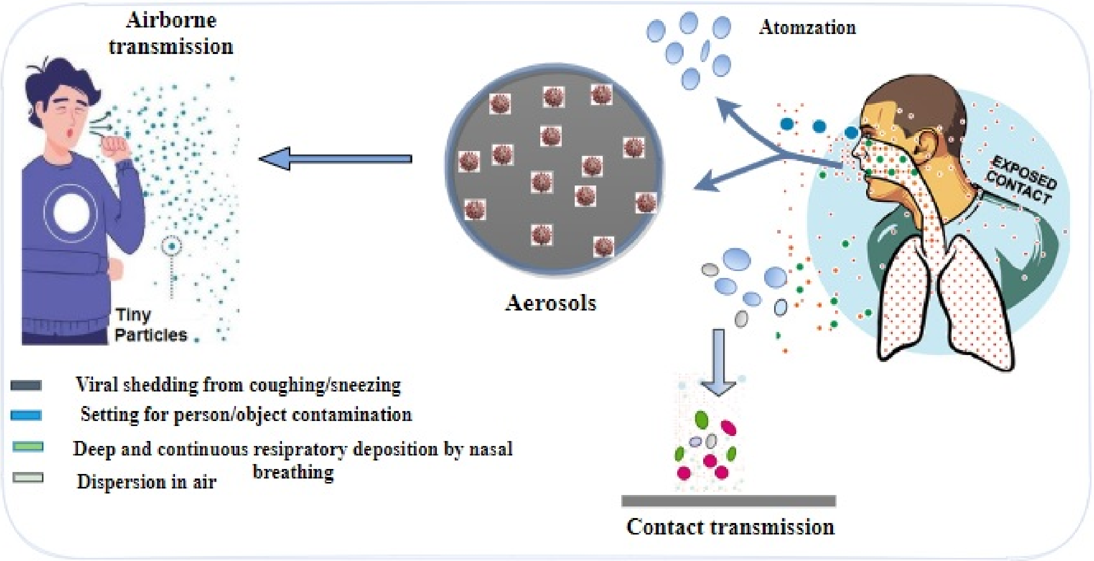


Figure 1.1: Mode of Transmission of Airborne Diseases

diseases do not cause by air pollution or any gases. ABD is one of the infectious disease which spreads rapidly by tiny pathogens in the in the air and transmitted directly or indirectly [4][5]. Figure 1.1 shown the different possible transmission media of the airborne diseases. The direct mode of transmission, particularly over short distance due to massive infections, aerosol that can be generated [6]and released by human expiratory actions like laughing, coughing, singing, and sneezing [7]. Human-to-human transmission of disease can result from direct contact with an infected person or an indirect contact through an intermediate object [8][9]. In 2003, the world-wide epidemic of a progressive risk of releasing anthrax or small-pox, H1N1 influenza epidemic in 2011, and Middle East Respiratory Syndrome (MERS) in 2013, has reiterated the potential serious threat to airborne infection to human health [10]. ABD has been discovered in humans for a long time. In North East Africa, smallpox epidemics have been reported [11].

In addition, increase of ABD people causes both natural and human activities to increase it. Some ABD like Tuberculosis, Coronavirus, Influenza, and Encephalitis requires continuous monitoring so that it spreads can be prevented on time [12].

Figure 1.2 depicts the classification of the infectious diseases [13]. During the research in the healthcare area, especially for monitoring and preventing the ABD at the initial stage can provide better health quality to human lives. The ABD increases, due to the rapid growth of industrialization which has become one of the cause to increase the pollution level in the air by releasing the harmful gases, dust particles which directly or indirectly affects the human health [14]. While the virus is mainly concern with public health, ABD are the major challengeable threat as an environmental issue [15]. Recently the pandemic of the coronavirus has put alert all around the world. Viruses like COV-2, influenza transmits people-to-people when they contact with each other and probably spreads when an infected person sneezes or coughs through the air pathogens [16]. These diseases badly affect socially and economically in concern to all over the world. ML based models are used to improve the performance in diagnosing ABD and the accuracy of prediction for screening both non-communicable and communicable diseases[13][14]. On November 11, 2021, the omicron is recognized as a variant of coronavirus (CorV) and its first case was detected in Botswana [17]. The omicron(OMCRN)cases have also been found in the United Kingdom, Italy, and Belgium. A high alert is announced

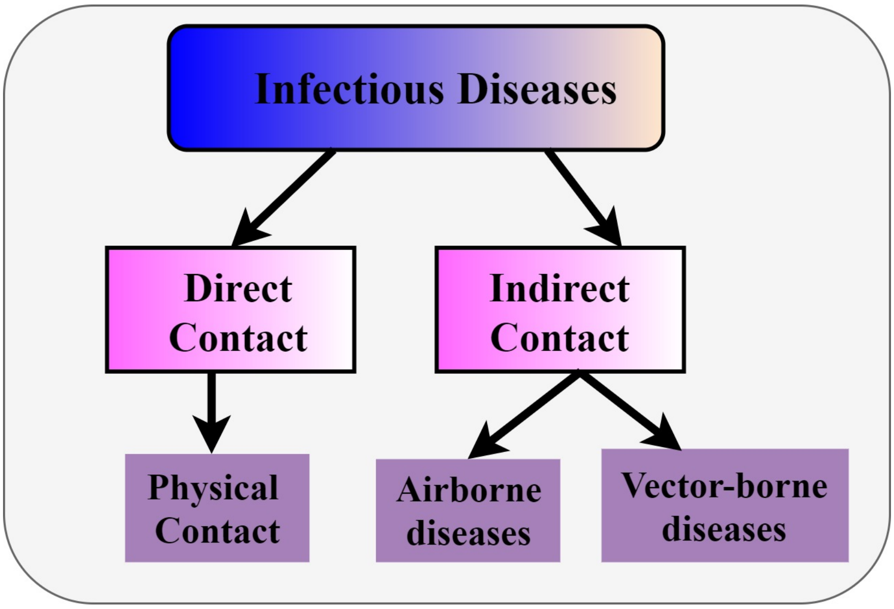


Figure 1.2: Classification of Infectious Disease

in Europe as another sequenced case has been found in Hong Kong and South Africa. The OMCRN is a heavily CorV variant also known as B.1.1.529 and is now listed by the World Health Organization (WHO) as a variation of concern.

1.2 Background

In 2002-2003, severe acute respiratory syndrome (SARS) spread into approximate 38 countries and killed more than 800 people. In 2009-2010, influenza A (H1N1) caused 17,000 deaths worldwide, whom were healthy people aged between 16 to 18 years. Avian flu incidence and calamitous events remain a serious threat to public health [18]. In 2019, 87% of TB cases was the leading in results of caused 1.2 million deaths [11]. In 2019, an outbreak of Acute Encephalitis Syndrome (AES) occurred in the northern region of the Bihar state of India [19]. The symptoms of AES are similar to West Nile, Japanese virus, and Nipah viruses. In 2019, an outbreak of Encephalitis occurred in the northern regions of the Bihar State of India. The WHO declared the pandemic AES is a neurological disorder disease that affects the spinal cord and the central nervous system of the human body when a virus attacks the body [20]. In 2020, the pandemic of coronavirus (CorV) had spread all over the world [21]. CorV is a viral disease that results in ill effects on humans and is recognized as public health concern globally [22]. In 2021, a report declared coronavirus as an airborne diseases. The objective of the current research is to diagnose and prevent the spread of airborne disease such as CorV. Table 3.1 depicts the major pandemics with their time-span. Every year number of people are affected by AB disease. The mortality rate is increasing day by day due to the spread of ABDs. It is a big challenge for the healthcare industry, caretakers, and government agencies to diagnose and prevent ABDs at an early stage. Therefore, the main objective of this research work is to design a novel framework to monitor and prevent the spread of ABDs viruses.

Table 1.1: Major Outbreak/Pandemics With Their Time-span

Pandemic Name	**Time-span**	**Type**	**Mortality toll**
Smallpox	1520-onward	Variola	55M
Flu	2000-2011	H1N1 Virus	20,0000
SARS	2002-2003	Virus	770
MERS	2015	Virus	850
AES	2019	Virus	More than 7000
SARS-CoV-2	2019-onwards	Coronavirus	More than 5.1M

1.3 Types of Airborne Diseases

In this section, Figure 1.4 depicts the history of past pandemics and their time span[1]. Table 1.2 describes the various category of diseases and respective description [23]. Some airborne diseases and their transmission media are discussed ahead:

Table 1.2: Various Category of Diseases and Their Respective Description

Category	**Description**
Air-borne Disease	Inhalation of pathogens including droplets or nuclei transmission causes damage to the respiratory system.
Diseases communicated via contact	Such diseases spread primarily via direct and indirect communication.
Diseases associated with Healthcare	Infectious diseases communicated while hospitalization or through healthcare services, or via various types of surgical operations.
Others	Risks associated with illness in general or a particular form of illness contracted.

(a) COVID-19

From 2020, millions of infections and thousands of global deaths are caused

---

[1] source:https://www.visualcapitalist.com/history-of-pandemics-deadliest/

by rapidly spreading a virus severe acute respiratory syndrome coronavirus (SARS-CoV-2) which belongs to the family of Coronaviruses, and it causes the disease COVID-19 [9]. When a person coughs or sneezes, this virus is spread through breathing droplets, which are largely considered airborne. Its symptoms include cough, fever, shortness of breath, tiredness, and loss of taste.

(b) Diphtheria

Diphtheria is one of the biggest causes of sickness and death mostly found in children of age group below five years. This disease damages the breathing system and can damage the heart, and nerves. In the last decade, very few cases have been reported in the USA because of the widespread vaccination. Antiviral medicines are the treatment for common diseases in the air. There is no treatment for certain diseases such as chickenpox but it is curable. Some medicines and other supporting treatments may aid in alleviating symptoms. Therapy for children with whooping toxin involves antibiotics and often requires hospitalization. Diphtheria may be treated with antitoxins and antibiotics if it may be seen in children.

(c) Tuberculosis(TB)

TB is transmitted by the air from an individual to another person. When people coughing, sneezing or spitting with lung TB, they spread the TB germs into the air. Only a few of these germs must be inhaled to get infected. In most productive years, TB affects adults. However, all age groups are at risk [24]. In 2019, 87% of TB cases was the leading cause of mortality from an infectious disease caused 1.2 million deaths [25].

(d) Influenza

It communicates very quickly because it is contagious. Since it is infectious a day before the signs of the disease appeara day before you notice the symptoms of the disease. Since it is infectious a day before the signs of the disease appear.

It remains for 5 to 7 days [26]. There are many strains to develop the flu [27]. Figure 1.3 depicts the exposure of the influenza viruses.

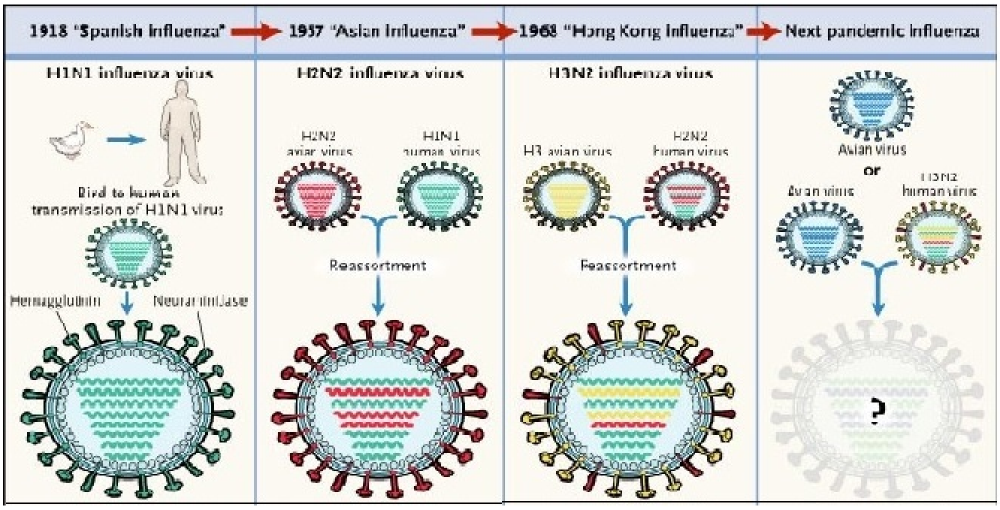


FIGURE 1.3: Influenza Viruses

(e) Measles

Measles, particularly in crowded conditions, is a viral disease and very contagious disease [28]. Most people just get measles once in their lives. It is a leading cause of death among children around the world and accountably led to 140,000 deaths in 2018. Around 23 million deaths from 2000 to 2018 are prevented by the measles vaccine. In the United States, the disease is less common and occurs in the most commonly unvaccinated people. In 2019, there have been 1,282 cases. Twelve confirmed cases were reported on 2 March, 2020.

(f) Whooping cough(Pertussis)

Whooping cough is a respiratory illness. It causes swelling in the airways which results in a persistent hacking cough. Every year, about 24.1 million cases occur globally resulting in 1,60,700 deaths [29].

(g) Chickenpox

The virus, varicella-zoster, causes chickenpox. The development of the disease after exposure will take up to 21 days. Most of the people get chickenpox only once in the life.

Therefore, global dialogue for diagnosing, forecasting, and predicting the ABD, the various themes are taken into account.

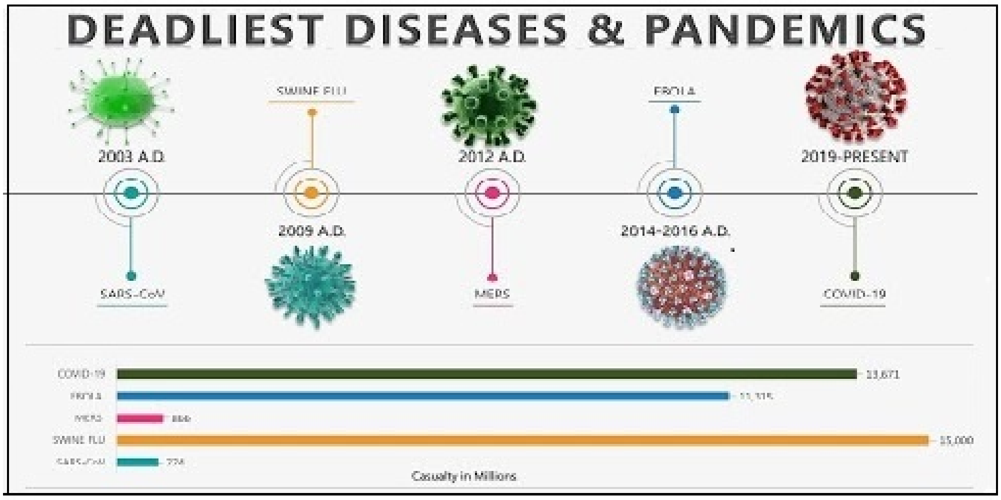


Figure 1.4: History oF Past pandemics

1.4 Common Symptoms

The common symptoms of ABD are:

- High viral fever,
- Headache,
- Sneezing,
- Coughing,
- Speech or hearing problems,
- Muscle pain,

- Vomiting,
- Damage respiratory system,
- Drowsiness, Tremors, Nausea,
- Unconsciousness, and
- Loss of memory.

Thus, to control and prevent such types of ABD, a strongly and smart healthcare system is required. In addition, increase of ABD people causes both natural and human activities to increase it, rapid industrialization has also caused the pollution level in the air, by releasing the harmful gases, dust particles into the atmosphere badly affected the human life. With advancement in technology as a root, machine learning algorithms may be applicable in the field of healthcare industry to early diagnose the disease with accurately and efficiently. Machine learning based framework incorporates to spatio-temporal data mining can help for controlling, monitoring and preventing the ABDs on time [30]. Henceforth, this framework helps to monitor disease continuously with accuracy, able to provide several healthcare services which were difficult with under developed technology. Therefore, this study aims to designs a novel based framework to control predict and prevent for the spread of ABD virus and its effects.

1.5 Delimitation of the Study

Airborne diseases are a vast area for research and several diseases are covered in this domain. Numerous researchers have endeavoured to find a suitable solution for these diseases covering all the possible areas for research. Due to a lack of time and inadequacy of financial resources, this study has been delimited to coronavirus and its variant "Omicron" out of various other airborne diseases like Tuberculosis, Chickenpox and Diphtheria. Further, this research can be used as the prototype for other airborne diseases like Tuberculosis, Measles, and Diphtheria etc.

1.6 Problem Identification

Airborne diseases, such as Tuberculosis (TB), Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), and Avian Influenza, pose significant challenges to public health globally, spreading rapidly and impacting a large volume of individuals. The identification and early detection of these diseases have become paramount concerns for both the healthcare industry and government agencies. With advancements in healthcare technology, there is a growing need for efficient monitoring and prevention strategies to curb the spread of airborne diseases while minimizing costs. Although various tests are available for detection, the regular screening for viral infections remains cumbersome. In 2020, the pandemic of COVID-19 underscored the severe impact of airborne diseases on a global scale. These diseases spread through respiratory secretions, such as coughing, sneezing, or spitting, which can linger in the air or settle on surfaces, posing risks of transmission through inhalation or contact with mucous membranes. Respiratory inflammation, caused by airborne bacteria, pathogens, or fungi, can lead to severe health complications, affecting not only humans but also non-human. Therefore, effective strategies for early detection, prevention, and management of airborne diseases are imperative to safeguard public health and mitigate the spread of infections.

The reason behind this research is to diagnose the disease on time from which the person is suffering. Due to an increasing number of patients day-by-day, it is a big challenge for the healthcare system to diagnose the disease at the initial stage and to control its spread. The proposed framework described in the research work serves a crucial purpose in monitoring the health of individuals and preventing the spread of airborne diseases. The presented framework incorporates various technological components and methodologies to achieve its objectives effectively. The main significant of the presented framework are:

(i) **Health Monitoring System:** The framework likely includes a sophisticated health monitoring system that collects data from various sources. These data can be collected from vital signs, activity levels, and any reported symptoms.

(ii) **Data Analysis and AI:** To make sense of the collected data, the framework may leverage data analysis techniques and artificial intelligence algorithms. These algorithms can identify patterns, anomalies, and early indicators of potential health issues. Machine learning models can be trained to recognize deviations from normal health parameters and prevent the spread of disease transmission.

(iii) **Real-time Alerts and Notifications:** Upon detecting any concerning signs or symptoms, the framework can generate real-time alerts and notifications. These messages can be sent to individuals, healthcare providers, or public health authorities, prompting timely action to address potential health risks.

(iv) **Contact Tracing and Epidemiological Analysis:** In the context of preventing disease spread, the framework may incorporate contact tracing capabilities. By analyzing movement patterns and interactions with other individuals, the system can identify potential exposure to contagious diseases. This information is invaluable for conducting epidemiological analyses and implementing targeted interventions to contain outbreaks.

(v) **Privacy and Security Measures:** Given the sensitive nature of health data, the framework must prioritize privacy and security. Robust encryption techniques and strict access controls can be implemented to safeguard individuals' privacy while still enabling effective health monitoring and disease prevention efforts.

(vi) **Scalability and Adaptability:** A successful framework should be scalable and adaptable to different populations, healthcare settings, and disease contexts. It should accommodate changes in technology, emerging infectious diseases, and evolving public health guidelines, ensuring its relevance and effectiveness over time.

In digital technology world, machine learning has emerged in the numerous domain of fields and capable of tackling issues in the era of agriculture, medicine, finance, economic and many more. Deep learning and Machine Learning approaches may be applicable in the field of healthcare industry to early diagnose disease with accurately and efficiently.

1.7 Machine Learning(ML)

Airborne diseases pose significant public health challenges worldwide, necessitating proactive measures for prediction and control. It is a sub-part of AI that main focuses to develop models and algorithms which are capable learning from and utilized that information to take decision and prediction based on data [31]. In recent years, this field has gained significant progress in many fields including healthcare and medicine due to its ability to extract valuable features and take decision making process efficiently [32]. ML has the ability of computers to learn without being explicitly programmed. In ML, a function, model, or algorithm is created to learn and perform extractions from existing datasets known as training datasets. Supervised learning is used to extract future events from training datasets when past data is available. ML algorithms offer powerful tools for analyzing complex data sets and making predictions, thus aiding in the prevention and management of airborne diseases. ML based models are used to improve the performance in diagnosing ABD and the accuracy of prediction for screening both non-communicable and communicable.

By leveraging machine learning prediction methods, healthcare providers can improve diagnostic accuracy, optimize resource allocation, and enhance patient outcomes by identifying at-risk individuals and intervening early in disease progression.

1.7.1 Machine Learning Approaches

ML consist three types of approaches i) Supervised, (ii)Unsupervised, and (iii) Reinforcement techniques depicted in Figure 1.5.

(i) **Supervised Learning** - In supervised learning, the algorithm is trained on a labeled dataset, meaning that each input is associated with the correct output. The goal is to learn a mapping from inputs to outputs, allowing the algorithm to make predictions on unseen data. Common supervised learning algorithms include:

- **Regression** - Used for predicting continuous outcomes. Examples include linear regression, polynomial regression, and support vector regression.
- **Classification** - Used for predicting categorical outcomes. Examples include logistic regression, decision trees, random forests, and support vector machine (SVM).

(ii) **Unsupervised Learning** - In unsupervised learning, use unlabeled input data to discover hidden patterns. Common unsupervised learning algorithms include:

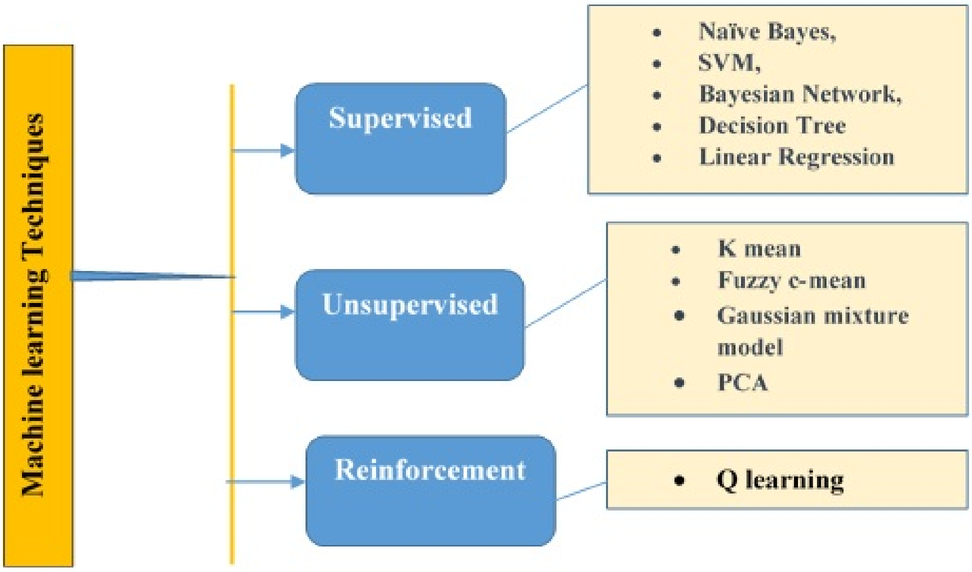


FIGURE 1.5: Classification of Machine Learning

- **Clustering**- Grouping similar data points together. Examples include K-means clustering, hierarchical clustering, and DBSCAN.
- **Dimensionality Reduction** - Reducing the number of features or variables in the dataset while preserving its important structure. Examples include Principal Component Analysis (PCA) and Stochastic Gradient (SC).

(iii) **Reinforcement Learning** - It consist combination of both supervised learning (Labeled data) as well as Unsupervised learning (Without labeled data). Examples include Q-learning, deep Q-networks (DQN), and gradient methods.

By leveraging various data sources, including environmental factors, population demographics, and disease transmission dynamics, ML algorithms can enhance our understanding of airborne disease spread and facilitate early detection and intervention strategies [33] [34]. Figure 1.6 depicts the various category of machine learning algorithms for airborne disease prediction based on Spatio-temporal data

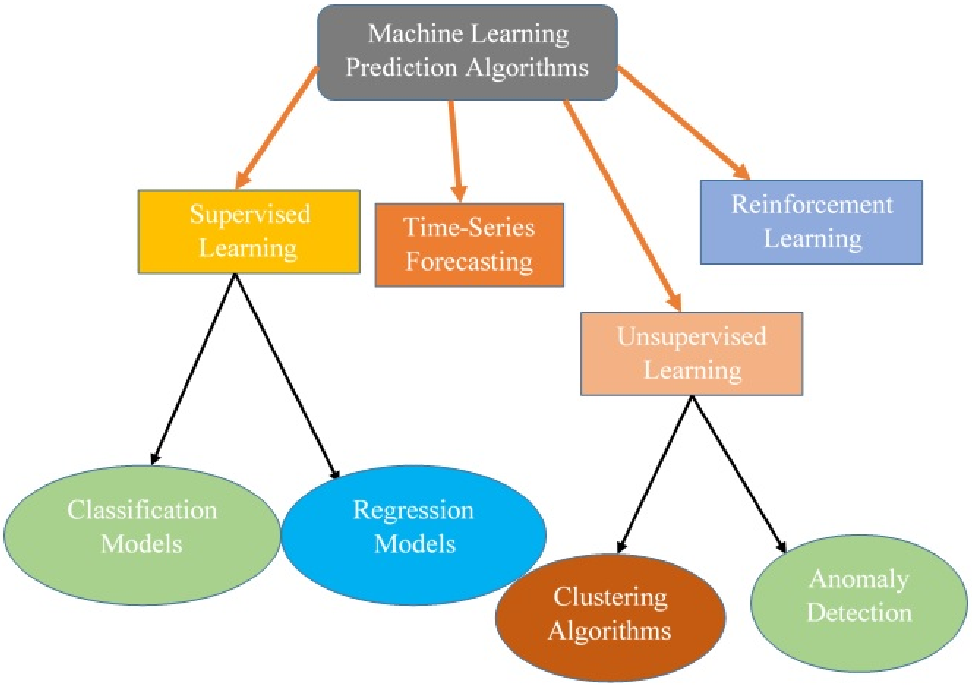


FIGURE 1.6: Hierarchy of Machine Learning Prediction Algorithms

mining. Various Machine Learning Prediction Algorithms for Airborne Diseases are discussed ahead:

1. **Supervised Learning Algorithms**

   - **Classification Models:** Algorithms such as Support Vector Machines (SVM), Random Forests, and Neural Networks can classify regions or populations based on their susceptibility to airborne diseases. These models can incorporate features like air quality, climate data, population density, and previous disease outbreaks to predict disease prevalence or risk.
   - **Regression Models:** Regression techniques like Linear Regression, Decision Trees, and Gradient Boosting can forecast the spread of airborne diseases over time. By analyzing historical disease data alongside

environmental and demographic factors, these models can provide insights into future disease trends and inform resource allocation and intervention strategies.

2. **Unsupervised Learning Algorithms**

   - **Clustering Algorithms:** Unsupervised techniques such as K-means clustering and Hierarchical clustering can identify patterns and clusters within airborne disease data. By grouping similar regions or cases together based on disease characteristics and environmental factors, these algorithms can help identify high-risk areas for targeted intervention and surveillance.
   - **Anomaly Detection:** Anomaly detection algorithms like Isolation Forest and One-Class SVM can detect unusual or unexpected patterns in disease transmission data. By flagging outliers or anomalies in disease spread, these algorithms can alert public health authorities to emerging threats or unusual transmission dynamics, facilitating prompt investigation and response.

3. **Time Series Forecasting**

   - Time series forecasting techniques, including Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) networks, can predict future disease incidence based on historical data. By capturing temporal patterns and geographically data in disease transmission, these models can provide short-term and long-term forecasts to support proactive public health measures and resource planning.

4. **Reinforcement Learning**

   - Reinforcement learning algorithms, such as Q-learning and Deep Q-Networks, can optimize decision-making strategies for disease control and intervention. By simulating disease transmission dynamics and assessing the impact of different control measures, these algorithms can recommend adaptive policies to minimize disease spread and the burden on healthcare system.

Table 1.3 summarize the numerous ML algorithms based on their approaches. The ML approaches such as clustering, classification and others, incorporates by using Spatio-temporal techniques can be used to extract meaningful information and for diagnosing ABDs [35]. The missing values and normalization on the experimental data are done by using data mining techniques [36]. The various ML algorithms are used for the classification of different ABDs. The Spatio-temporal based data mining techniques such as Recurrent Neural Network and Convolutional Neural Network have been used for the prediction of risk factors of ABDs such as Coronavirus, Tuberculosis, Encephalitis, Mumps, and others [37].

1.7.2 Various ML and DL algorithms

Various machine learning and deep learning algorithms are discussed below:

Support Vector Machine (SVM)

It is a supervised machine learning algorithm. SVM is a non-stochastic binary linear classification technique that is utilized for the categorization of the input data [38]. SVM is also used to determine the regression problems. It helps to divide the data by producing a line called a hyperplane. Support Vectors are the data points in each class that are nearest to the hyperplane, as determined by the SVM algorithm [39]. The separation is shown by the hyperplane from the support vectors so that the hyperplane is visible through them. The new data point will be consisting either one of the two classes [40].

Random Forest Regression (RFR)

The RFR approach is utilized for the decision-making problems that generate various decisions from an input dataset. Such an algorithm divides the datasets into

Table 1.3: Various Machine Leaning Algorithms

Approaches	**Methods**
• Bayesian	• Naive Bayes • Gaussian Naive Bayes • Bayesian Network • Bayesian Belief Network
• Clustering	• K-Means • Hierarchical Clustering • Expectation maximization • Probability Clustering
• Decision Tree	• Classification and Regression Tree • Conditional Decision Tree • C 5.0
• Deep Learning	• Convolutional Neural Network • Recurrent Neural Network • Long Short Term Memory Networks • Deep Belief Network
• Dimensionality Reduction	• Principal Component Analysis • Stochastic Gradient
• Neural Network	• Perceptron • Convulational Neural Network • Recurrent Neural Network
• Regression	• Linear Regression • Logistic Regression
• Ensemble	• Randon Forest • Boosting • AdaBoost • Grandient Boosting Regression Tree • Bootstrapped Aggregation(Bagging)

various sub-parts before creating further sub-trees. The final output is created by combining each decision's regression tree which helps to predict a forecast and produce more efficient and accurate outcomes. The output value of each input in RFRR is the average of all values predicted by several decision trees. The problem of overfitting is not handled by a decision tree algorithm as effectively as a decision tree. There are several causes, but the most frequent one is overfitting caused by each tree starting to build rules for the described class. The generalized outcome will be more reliable as it gets a vote. However, a new data value is calculated by mean, the predictions of the decision trees [41].

K-Nearest Neighbors (KNN)

The KNN algorithm is used for both classification and regression purposes [42]. To classify new data, this algorithm finds the distance between it and all other points in the dataset. It often uses Euclidean distance to find the distance between two points. It then selects the k nearest neighbours and determines the predicted class for the predicted value by averaging their target values [43]. The limitation of kNN, it can be computationally expensive for large datasets and may not perform well in classifications.

J48

A decision tree is used to extract the training dataset while taking a number of characteristics and goals into consideration. It is the application of an algorithm for classifying and determining data according to the attributes of the given data. It is a decision tree that considers several qualities and ideals to obtain data from the training set. It is the application of an algorithm to the classification and determination of data according to the values of its attributes[44].

Artificial Neural Networks(ANN)

ANNs is used in various fields such as pattern recognition, regression, and classification. Basically, it is utilized in numerous applications of various domains including image and speech recognition, natural language processing, and predictive modeling. Comprising interconnected neurons and input layers. It processes the information through weighted connections and activation functions of the network. Input data is fed into the network, propagated through hidden layers, and then output is eventually produced. During training, ANNs adjust the connection weights based on the input data and expected output, typically using algorithms like backpropagation.

Long Short-Term Memory(LSTM)

LSTM is a type of recurrent neural network architecture that is designed to resolve the vanishing gradient problem associated with recurrent neural networks. LSTM is capable of storing long-term dependencies in data by maintaining a memory cell that can store information for a long period of time. LSTM can be effectively used in speech recognition, language modelling, and time series prediction. It consists of various gates namely input gate, forget, and output to control the flow of information into and out of the memory cell. It allows forget cells to selectively remember or forget certain pieces of information. This mechanism enables LSTM to learn long-term data dependencies and make to more accurate predictions.

1.8 Spatio-temporal Data Mining

Spatio-temporal (ST) applications are domains where data is collected on different spatial (geospatial) locations at different points in time. It's prediction consists in trying to forecast the future value of the target variable for some location, based on information on past values of this variable at that same location as well as on nearby locations [45]. ST based prediction in healthcare involves using both spatial (Geographical-based) and temporal (time-based) data to forecast various health-related outcomes [46]. This approach leverages data from diverse sources such as e- records, medical imaging, wearable devices, environmental factors, and socioeconomic indicators to make predictions about diseases, outbreaks, patient outcomes, and more. The following are the key implications of ST-based prediction:

(i) **Data Collection-** Relevant data is collected from various sources including hospitals, clinics, wearable devices, environmental sensors, and public health records. This data includes information such as patient symptoms-based data, medical history, geographic locations, time stamps, environmental factors, and many more.

(ii) **Data Integration-** As data are collected from different types of resources so such heterogeneous data are integrated into a unified framework by utilizing various techniques such as data normalization and cleaning. Spatial data may include geographic or regional coordinates while temporal data includes timestamps, dates, or time intervals. Temporal data may be collected at different time stamps from different locations [30].

(iii) **Feature Extraction-** Features are extracted from the integrated data so that it can be used for prediction the disease in healthcare system. ST based data mining techniques may be used for extracting time -spatial features such as trends in patient vital signs over time [47].

Data segments and their temporal and geographical relationships are made possible via time-series data extraction patterns [48].

(iv) **Model Development-** Machine learning and statistical models are trained using the extracted features based on spatial-temporal data to make predictions. These models can range from traditional statistical methods like linear regression to more complex techniques such as neural networks, spatio-temporal models like spatio-temporal autoregressive models, or recurrent neural networks (RNNs).

(v) **Prediction and Evaluation-** The trained model is used to predict health-related outcomes based on spatial-temporal pattern data. The predictions are then evaluated by using statistical metrics such as accuracy, precision, recall, f-measure, or area under the curve (AUC).

(vi) **Decision Support-** The predictions generated by the model can inform healthcare providers, caretakers, and stakeholders so that proper actions can be taken promptly.

Moreover, Spatio-temporal based prediction in healthcare can provide potential for monitoring, improving patient outcomes and diagnosing the diseases accurately and efficiently through the integration of spatial-temporal data analytics.

1.9 Aim of the Research

According to World Health Organization(WHO), airborne diseases have many harmful impact on the individual life's. Recently in the past few years, Coronavirus has been perceived as a worldwide pandemic and badly affected all across the world. Although the COVID-19 cases have been decreased but still it has become remained a significant global health concern. The proposed work provides an efficient framework for monitoring the ABD over a spatiotemporal basis. The outbreak of ABD and its spread to the nearest regions is a big preventing challenge for healthcare and government agencies. The proposed framework helps in monitoring and preventing the spread of ABDs with accuracy and an efficient manner.

1.10 Objectives

(i) To analyze the prediction performance of existing frameworks.

(ii) To develop an effective prediction methodology for ABD prediction over a Spatio-temporal.

(iii) To Validate the predictive performance of the proposed methodology with state-of-the-art prediction models.

1.11 Research Contribution

The inefficiency of traditional methods for diagnosing the disease with some obsolete methods has led to an increasing in the mortality rate. Due to inadequate and inaccurate data, does not provide any help to identify the disease accurately causing the inaccurate diagnosis of the disease in results. AI-based technology has already seen several uses in the modern period, with significant advancements in the field of healthcare. Since its beginning, this technology has advanced quickly. ML-based DL technology has put their potential with significant advancements such as in the field of healthcare for several uses in the recent period. This technology has progressed quickly in various domains such as in the healthcare system for diagnosing diseases accurately and timely. The current study incorporates a machine-learning approach providing a novel framework for diagnosing and identifying airborne diseases. In this present work, DL which is a sub-domain of ML CNN-based prediction model has been utilized for the classification and identification of the coronavirus disease. The main contribution of this research is as follows:

1. A CNN-based SQueezeNet model is presented for the classification and diagnostic the Coronavirus disease at early stage.

2. To enhance the efficiency of proposed framework, the presented study focuses on binary and multi-class classification from three obtained datasets.
3. An augmentation is performed to the imbalance problem of the dataset. In addition, the advantage of transfer learning is applied to solve the problem of overfitting and speed of convergence.
4. To enhance the research work, the current work has also proposed a prediction model for monitoring, identifying, and preventing the spread of the omicron(variant of coronavirus).
5. For the classification accuracy, weighted naive based algorithm has used to determine the omicron severity index of an individual. to increase the predictivity efficiency, LSTM-ERNN based prediction model has been used for diagnosing and predicting the omicron.
6. To explore the validation of the entire system, the presented model has evaluated with various statistical parameters and errors.
7. The experimental results have shown the proposed system outperformed for coronavirus and omicron in terms of accuracy, specificity, recall, and f-measure.

1.12 Research Methodology

This section provide the step-by-step information that are utilized for diagnosing the ABDs. It is a process to describe the researcher work performed in this study. The flowchart of proposed methodology is depicted in Figure 1.7.

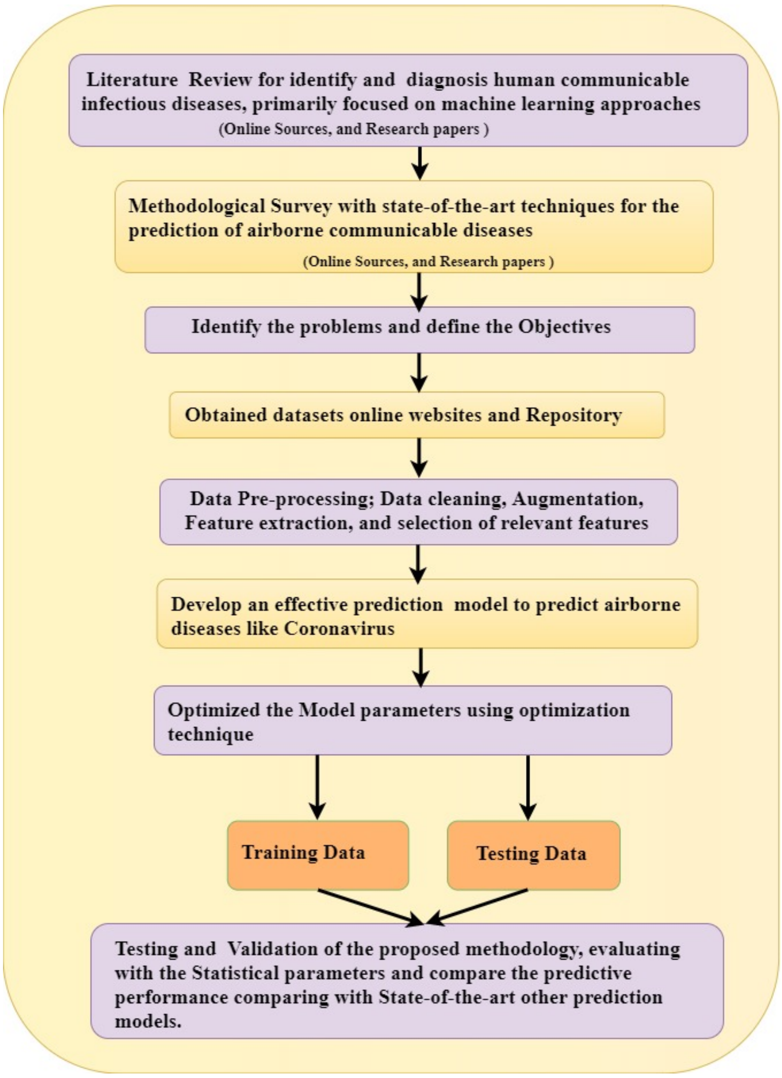


FIGURE 1.7: Workflow Diagram of Research Methodology

1. To accomplish the first objective, the researcher explored the various research papers, collected from the online, web of Science, and Google Scholar and focused on machine learning approaches for diagnosis and preventing the spread of communicable airborne diseases.
2. To achieve the second objective, the datasets were collected from UCI repository for identifying the coronavirus disease. In addition, data preprocessing in which data cleaning and augmentation process has been performed for analyzing the data. Furthermore, a novel framework for the monitoring and

prediction of airborne disease is developed.To illustrate the efficacy of the suggested model, it has assessed in comparison to other state-of-the-art approaches.

3. Finally to achieve third objective, the proposed model has been evaluated with numerous statistical parameters and outcomes are compared with the state-of-the-art other prediction models.

1.13 Motivation

With the advancement in technology embedded with ML which is sub-part of Artificial Intelligence has been successfully adopted by numerous healthcare industry to achieve certain tasks such as prognosis, monitoring, and diagnose the disease on time [49]. In this domain, continuous regular monitoring of an individual is not only a challenging area but also requires a real-time based application framework to control and prevent the spread of the disease. Some ABD like Tuberculosis, Influenza requires continuous monitoring so that it spreads can be prevented on time. During the research in the healthcare area, especially for monitoring and preventing the ABD at the initial stage can provide better health quality to human lives. By diagnosing the ABD disease, and predicting the spread of such infectious diseases are really a challenge in the real-time system. However, this work helps to provide an efficient solution for monitoring and predicting the spread of ABD in real-time, which badly affects human health.

1.14 Scope of the Study

ABD is a prospective application domain which badly affects human life regularly. In the digitalized era, health problems are increases day by day. By the growth of industrial pollution rapidly increasing the harmful dust particles and social interaction, and many more hazards increase the diseases at an alarming rate. Many countries have declared ABD disease as a major threat to human life, which badly affects human life. ABD may altogether restrict the limit of a person as per the everyday manner. Its frequency adversely impacts an individual's instructive and social fulfillment. The reason behind the study will address the critical challenges to predict the ABD, which are considerably more hazardous for a human being, Government and healthcare agencies.

1.15 Thesis Organization

This thesis is organized into six chapters as shown in Figure 1.8

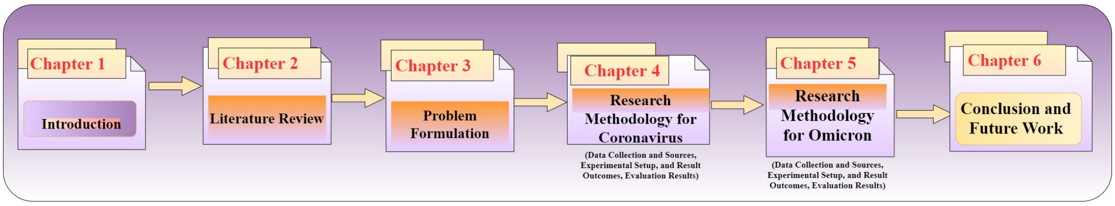


FIGURE 1.8: Chapter-wise Thesis Organization

Chapter 1 : Introduction This chapter introduce about the airborne diseases, its types, common symptoms, effects, and its communicable resources. Moreover, this chapter also describe the various machine learning techniques and methods spatio-temporal data mining that are pervasive in a healthcare system.

Chapter 2 : Literature Review The researcher can gain more context and a deeper understanding of the domain area with the help of this chapter. After addressing the many research gaps that exist in the pertinent field of study, it goes on to outline the most current algorithms that have been devised to solve the problem. Various AI existing techniques and machine learning prediction analysis approaches which are used by researchers in the current domain of healthcare system. This chapter is derived from the following paper:

- Sapna Kumari, Munish Bhatia “Machine Learning Techniques For Public Health System: A Scientometric Review”, 2022, 2nd International Conference on Computer Science, Engineering and Applications
- Sapna Kumari, Munish Bhatia “A Systematic Analysis on Airborne Infectious Virus Diseases: A Review”. ”1st International Conference on Innovations in Data Analytics (ICIDA 2022)“

Chapter 3 : Problem Formulation This chapter discussed the proposed methodology of the research works. The presented work consists two research works one for classification of coronavirus disease and second research work for Analysis, Monitoring and Identifying Omicron (Variant of Coronavirus).

- Sapna Kumari and Munish Bhatia “A cognitive framework based on deep neural network for classification of coronavirus disease” “Journal of Ambient Intelligence and Humanized Computing”, Springer 2022
- Sapna Kumari, Harpreet Kaur, Prabal Gupta entitled “A Cognitive Effective Framework for Analysis, Monitoring and Identifying Variant of Coronavirus” in ”The Journal of Supercomputing”, Springer 2024

Chapter 4 : Research Methodology For Coronavirus This chapter discussed the first research work, the proposed methodology for the identification of coronavirus. The numerous layers have been utilized by the presented framework to achieve the objectives described in this chapter. The detail of each layer and step-by-step process of the presented model is described in various subsections of this chapter. Moreover, this chapter also contains the experimental setup, dataset collection sources, data pre-processing, feature extraction, data and classification analysis techniques, and evaluated results and performance comparison for the research work have described in depth. This chapter is particularly derived from:

- Sapna Kumari and Munish Bhatia "A cognitive framework based on deep neural network for classification of coronavirus disease" "Journal of Ambient Intelligence and Humanized Computing", Springer 2022

Chapter 5 : Research Methodology For Omicron This chapter discussed the second research work, the methodology process for the identification of Omicron. The presented model consists layered architecture and the detail of each layer is described in various subsections of this chapter. Moreover, this chapter also describe the data collection source and showed the outcomes of the implemented model. The outcome of the presented model performance has been compared with other state-of-the-art models. This chapter is particularly derived from:

- Sapna Kumari, Harpreet Kaur, Prabal Gupta entitled "A Cognitive Effective Framework for Analysis, Monitoring and Identifying Variant of Coronavirus", in "The Journal of Supercomputing", Springer 2024

Chapter 6 : Conclusion and Future Scope This chapter contains the conclusion of the results, to accomplish the task of the research works. The chapter concludes the thesis with future study showing the direction to improve the suggested approach.

Chapter 2

Literature Review

This chapter shows the existing work done by various researchers. Many Research is on-going in this field, but still, many technical issues need to be resolved in order to accuracy, performance, and efficient solution for medical applications.

2.1 Literature Review Based on Infectious Diseases

[50] described that the presence of bioaerosols has been suspected as the cause of various human diseases covering not only infectious/respiratory which leads to cancer. Bioaerosols are also found in most 649 enclosed environments as various internal sources that are generally associated with human activities. To prevent the increase of aerosols cleaning and maintenance activities can play a vital role to increase indoor air quality .
Sareen et al. [51] proposed a framework to healthcare services using RFID technique for monitoring and controlling the outbreak of Ebola virus Patients. Further, the authors used temporal network analysis for analyzation and described the current state of the patients. The experimental results predict more accuracy of the

data for analyzing the outbreak.

Sood and Mahajan [52] developed a model to forecast the trajectory of a Chikungunya virus outbreak. Additionally, the authors provided a symptom-based analysis of the West Nile virus and Zika. Additionally, the scientists suggested a fuzzy method that uses a real-time application system to identify infected people and send a warning message to medical professionals.

Hassan et. al. [53] proposed a conceptual design for patient monitoring based on medical sensors to collect data for predicting and controlling the dengue outbreak. The author discussed the indoor and outdoor sensor parameters but not focused on architecture design .

James et al. [54] presented a descriptive study for workers and laypeople. By taking various cases of Airborne infections, authors made questionnaires in which ten questions for staff and twelve questions for laypeople had prepared. They concluded based on a survey that more awareness is required to staff and laypeople for preventing infections.

Shanthamallu et al. [55] introduced various machine learning and learning modalities, including supervised, unsupervised methods, and deep learning paradigms. Authors also discussed applications of machine learning algorithms in various fields, including pattern recognition, sensor networks, anomaly detection, Internet of Things (IoT), and health monitoring .

[56] proposed an IoT-Cloud based wireless technology for the analysis of seizures. Selected features are input to a k-means classifier to detect epileptic seizure states in a reasonable time. The author used the cloud for storage, processing after analyzing the data and by using the GPS technique to send alert to the mobile of the patient with efficiency and accuracy. The performance of the model was tested on Amazon EC2 cloud and performance was compared in terms of execution and accuracy.

Tuli et al. [57] introduced Healthfog, a cutting-edge fog-based smart healthcare solution that uses deep learning and IoT to diagnose cardiovascular diseases. In addition to managing patient health data that is effectively retrieved from a variety of sensor devices, the model that is being presented offers healthcare services. It is used with deep learning to create a real-time healthcare prediction application. In a fog computing environment, the system is monitored using the FogBus framework. The suggested system's efficiency is evaluated in terms of training, testing, and power consumption.

[58] discussed the definition of aerosols, consider the 'aerosol transmission' in the context of some infectious agents that are well-recognized to be transmissible via the airborne route. Further authors discussed the environmental factors and also explained the various prevention methods like the type of personal protective equipment (PPE) for intervening from the potential mode of transmission.

Sweileh et al. [59] discussed the various health-related literature" related to climate change and "infection-related literature". A bibliometric analysis was performed on 4247 documents and the results have shown Documents in infection-related literature had a higher h-index than documents in the health-related literature. The author concluded that more research on health especially in the direction of infectious disease is required .

Bhatia et al. [19] suggested a framework for real-time probabilistic vulnerability prediction based on health status analysis. Additionally, the authors conduct 14-day simulations data on a number of individuals with varying qualities in order to assess the applicability of the suggested framework.

Bhardwaj et al. [60] discussed the role of ML algorithms in enabling Iot based healthcare applications. Authors conducted a comparative study of ML classifiers for the detection of Influenza. This study focused on detecting the level of cardiovascular disease in young one by using a Naive base classifier. Furthermore, these algorithms can predict the spread of contagious disease effectively.

Sood and Mahajan [61] suggested a unique remote control method to identify and prevent the mosquitoes borne disease. The authors of the suggested work concentrated on early mosquito borne disease infection detection. Data extraction and collecting are done using a fast automated keyword extraction approach. Information is processed and stored in the fog-cloud region. Data are categorized as infectious or non-infectious using a decision tree.

Sukhralia et al. [62] investigated the invasion of arthropod-borne viruses that cause illness in public health. The main focus of this paper was the common families of mosquito-borne arthropod-borne viruses that pose a threat to mankind viruses including Zika virus, Japnese Encephalitis, and the Nile virus.

Sood et al. [63] proposed a novel framework for monitoring and determining the level of health. The primary purpose of this framework is to forecast the disease's severity. The authors' severity study was centred on the health of the students. The experiment's data gathering and classification techniques for identifying the health vulnerability were done using the UCI repository.

2.2 Scientometric Review

Scientometrics is the quantitative assessment of scientific and technical publications. In 1969,Mulchenko and Nailmoy were firstly developed the term scientometric [64]. It provides a statistical evaluation of the advancements in science and technology and is the cornerstone upon which future research goals must be chosen and addressed. It promotes the growth and diversification of the field of study [65]. Furthermore, it offers analytical metrics for assessing a study topics and related subfields quantitative scientific performance [66]. It includes observing publishing patterns, citation trends, keyword co-occurrence analysis, and different quantitative metrics from 2000-2020. This review analysis determined the quantitative aspects of the science using various bibliometric approaches to evaluate the development of research, Geograhical based analysis [67], social relevance towards the global health system [68]. This study explore from the web of science database from time span 2000-2021.

Google Scholar, Web of Science, Scopus are the three primary sources used for bibliometric studies [69]. In this review, data are collected from Web of science core collection due to its reliability. Several data visualisation tools, such as VOSviewer [70], CiteSpace [71], Gephi [56], and CiteNetExplorer are readily available [72]. VoS viewer which is a java based tool that consist visualization and mapping exploration, has used for visualization analysis of keyword co-occurence in the current study.

The database is searched for the query TI=(”airborne diseases” OR ”climate change and airborne diseases” OR ”communicable infectious disease ” OR ”global healthcare”) in advance search field [73]. A total of 11,304 literature data found in the domain of global health literature in this domain. Duplicate and other types of data including book chapters and conferences have been removed. According to the refined retrieve strategies, a total of 9,602 data records have been retrieved.

2.2.1 Findings and Discussion

Next sub-section described the analysis of various countries which are actually contributed for the global health specially for the infectious communicable disease, keywords and their co-occurence, and other bibliometric metrics.

2.2.2 Publications Growth and their Cumulative Publications

The analysis represents the growth pattern of publications over the years to reflect the knowledge in a particular field of study.

A total of 11,304 literature were retrieved including 9623 articles, 498 proceeding papers, and 1183 review papers, on the whole, the scientific production in the year 2000-2020. Among these articles, 410 were published in 2011, 457 in 2014, 509 in 2017, and 795 in 2020 publications related to climate change and airborne

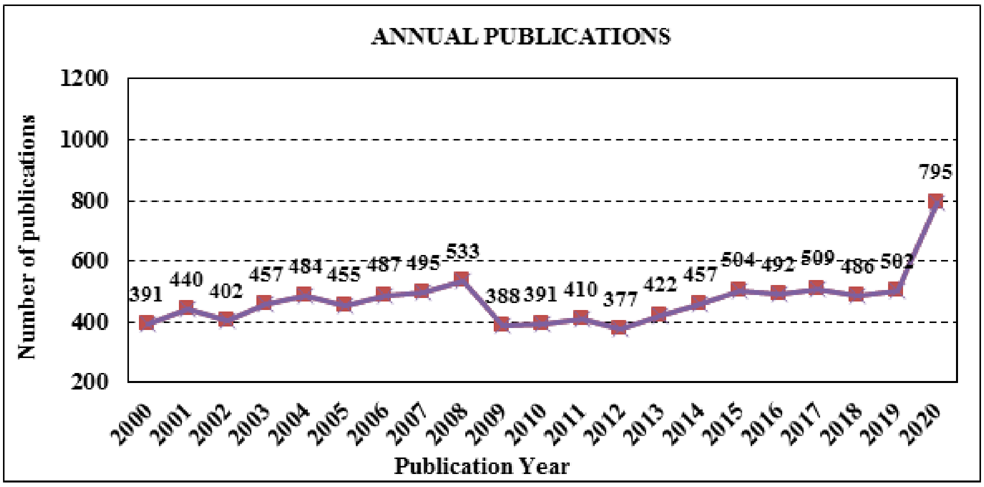


FIGURE 2.1: Annual Growth Rate

infectious disease.Table 2.1 shows the list of publications that were published year-on-year. Figure 2.1 shows the annual publication of every year in which it depicts that in 2008 publication share was 5.41%, in 2019 it increases 6.27%, and in the year 2020, the publication increased by 1.45%.

TABLE 2.1: Publications and their Growth Rate

Year	**Publications**	**In(%)**
2000	391	4.07
2001	440	4.58
2002	402	4.19
2003	457	4.76
2004	484	5.04
2005	455	4.74
2006	487	5.07
2007	495	4.82
2008	533	5.41
2009	388	4.04
2010	391	4.07
2011	410	4.27
2012	377	3.93
2013	422	4.40
2014	457	4.76
2015	504	5.25
2016	492	5.12
2017	509	5.30
2018	486	5.06
2019	502	6.27
2020	795	7.72

2.2.3 Geographical Distribution Analysis

The contribution of several countries in a certain field of study is represented through geographic distribution analysis. Table 2.2 shows the list of the top 20 active countries in publishing the documents towards the health and AIVD literature that has been dominated by European countries. However, the USA(America) led with 1832 (28.46%), followed by UK(England) with 1284 (19.95%). The ten countries had published the documents in the range of 200- 400 and eight countries published the documents 40-60 in the health field. Mapping research collaboration has shown that the USA (link strength=69) and England (link strength= 67) located in the middle of the map with the strongest link strength shown in Figure 2.2. For each country, the node side represents the document percentage with international researchers. Large node size indicative of greater collaboration with that country.

TABLE 2.2: List of Top 20 Productive Nation

Sr. no.	Country	Frequency	Documents (in %)	Citations	ACPD
1	USA	1832	28.46	138759	75.74
2	England	1284	19.95	73154	56.97
3	France	368	5.72	22639	61.52
4	Spain	324	5.03	17237	53.20
5	Canada	323	5.02	19229	59.53
6	Taiwan	272	4.23	13460	49.49
7	Italy	223	3.46	10570	47.40
8	Switzerland	220	3.42	15559	70.72
9	Belgium	164	2.55	8796	53.63
10	Australia	157	2.44	11331	72.17
11	Japan	135	2.10	7708	57.10
12	Germany	104	1.62	5839	56.14
13	Scotland	73	1.13	3949	54.10
14	South Africa	67	1.04	3713	55.42
15	Thailand	55	0.85	3197	58.13
16	Peoples R China	54	0.84	2778	51.44
17	Netherlands	52	0.81	3005	57.79
18	India	46	0.71	3130	68.04
19	Sweden	42	0.65	2326	55.38
20	Uganda	42	0.65	2502	59.57

ACPD - Average Citation Per Document

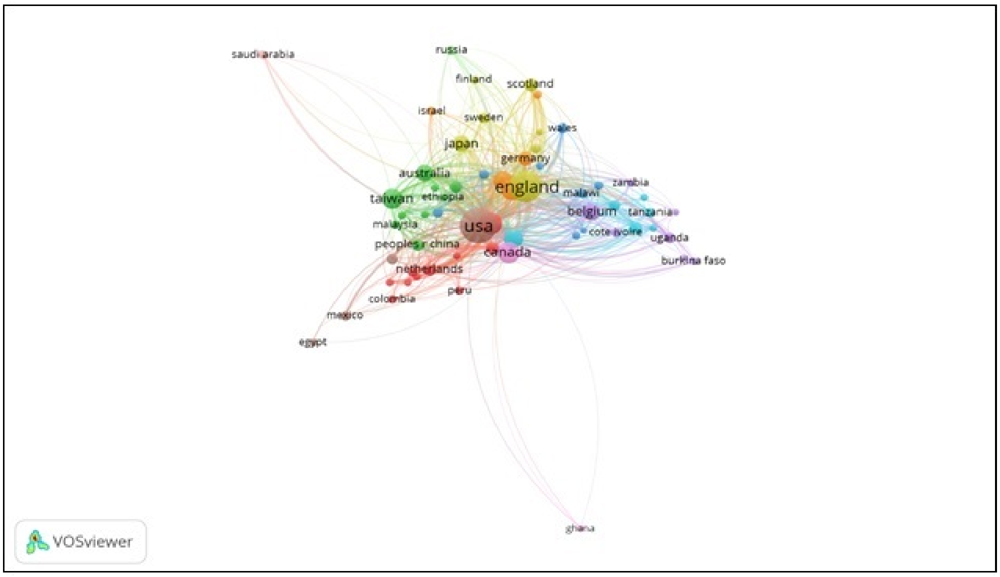


FIGURE 2.2: Visualization Map of Research Collaboration of Active Countries

2.2.4 Analysis of Keywords Co-occurence

Keywords has an significant role in scientific research. It abstractly defines the content present in scientific writing, represents the analyzing requirement of key topics, and aspects. It also helps to determine the research topics in a particular area. The current study retrieved 270 keywords from 4599 author publications having a minimum occurrence of 5 times in the past (2000-2020) years [74]. The VoSviewer takes only those keywords that match with the criteria of the minimum number of occurrences (MNK). Table 3 depicts that the total number of keywords (TNK) with a fixed value of MNK to get a co-occurrence network of selected keywords (SK). In Figure 2.3, the VoSviewer network map shows that each keyword represents by a node and the total Links between nodes are 2215. The bigger size of the node represents the higher frequency of a keyword. The less distance between two nodes indicates the stronger co-occurence between the two keywords. Table 2.3 shows the list of top 15 keywords that match with the criteria maximum number of occurrences five times in the publications. The keyword HIV (200) occurred the maximum number of times followed by epidemiology (129), and tuberculosis (104), and other keywords having in the range 20-70 to the author's

Table 2.3: List of the top 20 Keywords with their Occurrences

Id	Keyword	Occurrences	Total link strength
1	HIV	200	346
2	Epidemiology	129	249
3	Tuberculosis	104	166
4	Antiretroviral Therapy	63	105
5	Surveillance	56	96
6	Risk Factors	49	96
7	Outbreak	43	98
8	HIV Infection	43	62
9	Mortality	39	92
10	Aids	34	74
11	Diagnosis	34	67
12	Mrsa	33	39
13	Staphylococcus Aureus	32	50
14	Infection Control	31	48
15	Drug Resistance	30	45

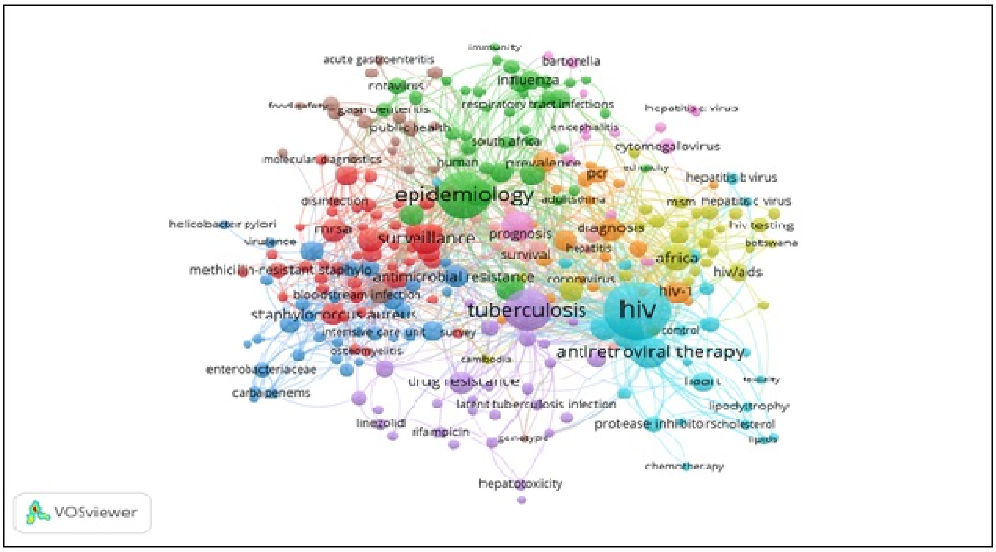


Figure 2.3: Co-occurence of Keywords

papers. The keywords that have the same meaning and plural of keywords are ignored during analysis in the network visualization of the map.

To minimize bias, find the research gaps, and synthesise the study various systematic literature reviews has been utilized in this thesis. There are numerous metods to perform systematic reviews [75]. For such purposes, It includes an organised review that focuses on commonly used methods, concepts and structures [64], Theme-based reviews [65][66], and a Meta-Analysis review [76], Bibliometric reviews [77], and Structured reviews [73].

2.3 Literature Review to Diagnose Diseases Using ML/DL Approaches

Nowadays, machine learning is a progressive study area that performs tasks without explicit programming. An algorithm, model, or function is developed in machine learning to execute extractions from pre-existing datasets[78]. ML is a branch of Artificial Intelligence that leverages statistical models and systems to learn from data in order to predict future sample attributes to complete specific tasks. DL which is a subset of machine learning that has significant approach for detection of disease through medical imaging [79]. As a result, the medical community has emphasized the need to focus on the advancement of diagnostic technology [80].

Nayak et al. [81] Eight convolutional neural network pre-trained models were compared by the authors. With an accuracy of 98.33%, the results show that ResNet-34 fared better than the other cutting-edge models in the classification of COVID-19 from typical occurrences. Recent studies have demonstrated CNN's capacity to use model performances to resolve difficult tasks including object identification, segmentation, and classification.

Jain et al. [82] used Chest x-ray and CT images deep learning techniques for diagnosing lung-related problems. Further, the authors compared 3 deep learning models with their performance.

Afshar et al. [83] suggested a framework model based on Capsule Networks for using X-ray images to diagnose Covid-19 disease. Several convolution layers and capsules are used to overcome the problem of class imbalance. Further, the authors have shown that COVID-CAPS performs satisfactorily on lesser trainable parameters in experimental investigation.

Ioannis et al. [84] employed an evolving neural network to automatically detect COVID-19. Specifically, a neural network has been used to apply the transfer learning process to the classification of COVID-19-induced pneumonia and common pneumonia, with an accuracy of 93.4%. Hemdan et al. [85] created a COVIDX-Net model that takes X-ray pictures into account. The COVIDX-Net model was trained using seven distinct CNN models, and it was verified using 50 X-ray images (25 normal and 25 COVID-19 cases).

Zhang et al. [86] offered a broad deep learning methodology for automatically identifying and evaluating regions that are highly susceptible to contracting COVID-19. The scientists performed a segmentation step using a DL-based approach in order to achieve this. After that, using predetermined measures, the contaminated regions in the CT scan were examined and measured.

Ramalingam et al. [87] proposed Heart disease prediction using machine learning techniques in which machine learning algorithms have been applied to various medical datasets to automate the analysis of large and complex data. The authors used various datasets for experimental results but the accuracy attained by model was not affected for each dataset.

Xu et al. [88] proposed a pre-trained Convolutional neural network to extract features using the CT images. These features are subsequently fed to next CNN layer to classify them into three classes. The experiment result has shown an overall accuracy of 86.7%.

Farooq and Hafeez [89] proposed a ResNet-based system for identification of CorV. The model's accuracy has remained constant at 96.2%.

Sood et al. [90] provides a scientometric review of the literature on COVID-19 study using ICT assistance. Artificial intelligence and medical imaging, mobile technology, ubiquitous computing, big data analytics, social media platforms, and printing technology are the six main categories into which ICT has been divided in this article. Using a variety of empirical techniques, including co-citation analysis, publishing and citation behaviour analysis, participating nations, and knowledge mapping of scientific literature using the visualisation tool CiteSpace, it thoroughly investigates the function of these technologies in COVID-19. In addition, it offers a visual method for locating research hotspots, evolving trends, developing paths, cluster analysis, and possible future directions in medical informatics.

2.4 Comparative Study Based on the Accuracy Using ML approaches

The majority of the papers focused on forecasting of recent pandemic disease coronavirus. Researchers has used the images data for diagnosing the CorV. Table 2.4 depicts the comparative study based on dataset and their accuracy.

2.5 Literature Review based on Medical Report Generation Using ML

many medical based analysis puts an important role in many research domain. Numerous researchers have contributed significantly for healthcare data analysis for real-time services and medical report generation systems [104] [105].
Chan et al. [103] described a model to detect pneumothorax in the small medical image. The authors used the SVM method to detect the symptoms of lung diseases using pattern analysis methods. The authors employed texture segmentation to represent the defective lungs in the proposed detection model. The proposed model was obtained higher accuracy as compared to other models.
Bai et al. [106] discussed the COVID-19 Intelligent Diagnosis and Treatment Assistant Program based on the Internet of Things. The main goal of the authors has to enable different levels of COVID-19 diagnosis and treatment by using medical technology.

Table 2.4: Comparative Literature study for diagnostic Coronavirus

Literature	**Datatype**	**Method**	**Findings**
kaya et al.[91]	Kaggle datasets (100 patients)	ResNet50	Accuracy:96.5%
Zhang et al. [92]	Github (905 Patients)	CNN and ML Algorithms	Accuracy:92.1%
Ozturk et al. [93]	Github (1000 patients)	CNN Model	Accuracy: 96.9%
Islam et al. [94]	Shenzhen Dataset	Ensemble	Accuracy: 91%
Wang et al. [95]	Repository (13045 patients)	Deep CNN Model	Accuracy: 98%
Alcantara et al. [96]	Local Dataset	CNN	Accuracy: 90.1%
Eudrado et al. [97]	Repository dataset (13,569 patients)	EfficientNet	Accuracy: 94.1%
Hernandez et al. [36]	Local dataset (800 patients)	ML algorithms	Accuracy: 97.12%
S. Ying et al.[98]	Dataset (200 patients)	ResNet-50	Accuracy: 86.1%
Li et al. [99]	Local dataSet (480 patients)	COVNet	Accuracy: 96.2%
Shi et al. [100]	Local dataset (1658 patients)	Random Forest method	Accuracy: 87.9%
MR Saybani et al. [101]	Local Hospital dataset (150 patients)	Support Vector Machine (SVM)	Accuracy: 98.33%
Tang et al. [102]	Dataset (176 patients)	RF-based model	Accuracy: 97.9%
J chen et al. [103]	Local dataset (255 patients)	Random Forest algorithm	Accuracy: 96.3%

Yildirim et al. [107] proposed a deep-transfer learning approach for automated diagnosis of diabetes mellitus. Dorj et al. [108] proposed an intelligent and rapid classification system of skin cancer using ECOC SVM, and a deep convolutional neural network.

In [109] discussed a computer-aided method that can help doctors categorize blood tumours, liver cancer, and several forms of breast cancer that are visible in photographs. The authors also intended to provide a framework for classifying these three types of disorders.

In the context of computer vision, For unsupervised re-identification of non-textual material, Zhou et al. [110] presented a multi-feature extraction using an adaptive graph learning model. To create a single learning model, the authors also include adaptive graph learning and multi-feature dictionary learning. Convergence is demonstrated using the optimisation technique. To demonstrate the superiority and efficacy of the suggested strategy, the experiment has been run on four datasets.

Alqudah et al. [111] used the CNN based model to detect COVID-19 infection by classifying chest X-ray images from normal, bacterial, and viral pneumonia cases.

Ucar et al. [112] proposed an efficient decision-making system for COVID-19. The authors used an AI-based structure design in addition to its light network that tuned for identification of COVID-19 with Bayesian network.

Li et al. [113] proposed an optimization algorithm is adopted to effectively solve the clustering problem by using one parameter required in the learning process.

Chang et al. [114] proposed a novel compound rank-k projection algorithm for bilinear analysis. In the proposed algorithm, the authors used the multiple rank-k projection models to find the best optimization solution.

Luo et al. [115] proposed an unsupervised feature selection to produce a faithful feature subset from the geometry structure of the original feature space by maintaining the intrinsic structure accurately. To achieve the subset, optimal reconstruction graph and selective matrix techniques are used simultaneously. Furthermore, the experiments are conducted over several data sets to verify the effectiveness of the proposed unsupervised feature selection algorithm.

Ren et al. [116] discussed the characteristics of the Neural Architecture Search as algorithms, its issues and also provided the comparative analysis for their research work in the paper.

Yu et al. [117] proposed the Adaptive Semi-supervised Feature Selection (ASFS) for cross-modal retrieval. In the proposed model, the authors used an efficient joint optimization algorithm to update the mapping matrices and label matrix for non-labeled data. Experimental results have shown the superiority of the proposed model.

In [118], the authors discussed adaptive optimum similarity matrix learning into the feature selection technique for chosing significant features. The authors deveoped a semi-supervised feature analysis framework for video semantic identification for the detection of disease.. Moreover, authors have been shown fine-tuned parameters with other methods to achieve higher accuracy.

2.6 Comparative Study For Diagnostic Coronavirus Based on DL models

Table 2.5 depicts the comparative study, strength and weakness for diagnosing coronavirus based on their models used in their research work.

2.7 Research Gaps

A research gap indicates a particular field which means something is never explored. The various researchers have worked in the field of medical but still more research is required to provide a healthy life to an individual. Many research and papers have mainly focus on trends and primitive topics but after all some points are never explored. In the field of Machine learning and Data Science Field, some research gaps are still exist as following:

Table 2.5: Strengthen and weakness comparison study of the selected Models

Authors	**Models**	**Strength**	**Weakness**
[119]	ncovnet	VGG-16 consist convolutional layers and five trainable dense layers for identification of CorV.	As the training loss value was still higher after the training is completed. The network was not trained until convergence.
[120]	ResNet50	Authors used ResNet-50 as the feature extractor and SVM for the classification	Authors uses few number of images for classification and not implemented the model for large datasets.
. [121]	CovXNet	multi-perceptron was used, with the group of convolution layers for classification	After 45 epochs, training convergence was unstable. The only fact is that each class has only 310 images, making overfitting easy to detect the coronavirus.
[122]	ResNet	To create synthetic data, it uses a deep convolutional generative adversarial network.	The study use the synthetic data and does not trained the network individually for each class.
[123]	COVID-Net	Authors has applied pooling approach to extract features of various scales	For COVID-19 cases, the experiment includes X-ray few images for the class detection performance.

1. Most of the researchers have focused only on improving the accuracy of the machine learning-based prediction model but not focused on other factors such as specificity, execution time, etc. The health, sector needs more research in medicine using AI and ML.

2. Majorities of the studies extensively explored the machine learning prediction models for diagnosing the disease but they did not discuss the prediction model techniques over the Spatio-temporal based data mining.

3. There is still a potential to work in the healthcare field where there is the urge to provide one common platform for the early detection, classification and identification of airborne diseases. The one platform will be helpful for accurately diagnosing airborne diseases.

4. Some obsolete technologies do not help diagnose diseases accurately. Due to the lack of resources in some rural areas, monitoring the patient is cost-effective. So, a cost-effective and efficient model can be helpful for monitoring and diagnosing patients timely. In addition, an alert generation message can sent to stakeholders if any abnormality is found so that any treatment can be taken promptly.
   Therefore, an efficient ML prediction model over the spatio-temporal technique will improve the accuracy of the ML models.

2.8 Conclusion

This chapter provides a systematic review that is reinforced by bibliometric research, and it is an important endeavor to provide the most comprehensive perspective on the evolving aspects towards the global health system especially related to airborne diseases. The main goal of this study is to provide insights to monitor, identify, and prevent the spread of airborne diseases using various machine learning approaches. This chapter insight and discusses the various existing ML approaches utilized in various types of research for diagnosing diseases. Many researchers have done much work to diagnose airborne disease but there are many limitations. In addition, this chapter explored research gaps relevant to existing approaches. Most of the researchers have focused only on improving the accuracy of the machine learning-based model but not focused on other factors such as execution time and statistical parameters NPV and PPV.

Chapter 3

Problem Formulation

3.1 Overview

Nowadays Airborne Diseases have sparked a big threat globally. These diseases are spread through air via tiny particles. Bacteria, fungi, or viruses spread these infectious diseases, but they are all transmitted through airborne contact by tiny pathogens. The direct mode of transmission, particularly over short distance due to massive infections aerosol that can be generated and released by human expiratory actions like laughing and sneezing. Such types of diseases are spread when an infected person sneezing, talking, coughing, laughing, or singing. ABDs can be communicable directly or indirectly [124]. ABDs like TB cause thousands of deaths per year, which are increasing in the mortality rate. Human contact and air droplets are the primary means of communication for the spreading of airborne illnesses. Human-to-human transmission can result from direct contact or indirect contact via an intermediate object or by touching the surface. Tuberculosis, and measles virus (rubella) require a long distance to transmit while whooping cough(Bordetella pertussis), Chickenpox(Zoster virus), Mumps virus, and Measles (rubella virus) are additional ABDs which can transmit over a short route distance (5m to 50m range) [125]. The current research work presented a hybrid framework for early diagnosing and preventing such types of diseases so that a proper treatment can be taken promptly if any abnormality occurred.

3.2 Background

In 2002-2003, severe acute respiratory syndrome (SARS) spread into approximate 38 countries and killed more than 800 people. In 2009-2010, influenza A (H1N1) caused 17,000 deaths worldwide, whom were healthy people aged between 16 to 18 years. Avian flu incidence and calamitous events remain a serious threat to public health [18]. The outbreak of encephalitis had affected the northern region of India. The symptoms of encephalitis are similar to West Nile, Japanese virus, and Nipah viruses. The WHO declared this outbreak as a neurological disorder disease that effects human spinal cord and the central nervous system when a virus attacks the body. Table 3.1 depicts the major pandemics with their time-span. Every year number of people are affected by AB disease. The mortality rate is increasing day by day due to the spread of ABDs. It is a big challenge for the healthcare industry, caretakers, and government agencies to diagnose and prevent ABDs at an early stage. Therefore, the main objective of this research work is to design a novel framework to monitor and prevent the spread of ABDs viruses.

Table 3.1: Major Outbreak/Pandemics With their Time-span

Pandemic Name	**Time-span**	**Type**	**Mortality toll**
Smallpox	1520-onward	Variola	55M
Flu	2000-2011	H1N1 Virus	20,0000
SARS	2002-2003	Virus	770
MERS	2015	Virus	850
AES	2019	Virus	More than 7000
SARS-CoV-2	2019-onwards	Coronavirus	More than 5.1M

3.3 Problem Identification

In the last few years, many people have lost their lives due to airborne diseases such as coronavirus. In 2020, the pandemic of coronavirus was spread to more than 185 countries. The novel coronavirus has investigated more than a billion of confirmed cases and thousands of people died. Moreover, In 2019, the outbreak of Acute Encephalitis Syndrome (AES) occurred in the northern region of Bihar (India). More than seven thousand people were killed due to this disease. Henceforth, a prediction model based on spatiotemporal information is utilized to manage medical resources in order to facilitate effective health-oriented decision-making and information delivery to an individual [126].

Nowadays, it is a very big challenge for the healthcare industry and for government agencies for accurate detection of the disease at the initial stage. In the current era of development, ML which is part of AI giving enormous services in the field of healthcare. Obsolete healthcare technology was not able to monitor and prevent the spread of such diseases efficiently. Many tests are available for detection, but it is cumbersome for viral detection regularly. The reason behind this research is to diagnose the disease on time from which the person is suffering. Due to an increasing number of patients day by day, it is critical to develop a model that could rapidly diagnose the disease at earlier and prevent its spread. ML which is a sub-part of artificial intelligence, has advanced rapidly, becoming an indispensable approach for resolving complex problems including speech recognition, computer vision, and natural language processing. Therefore, in the current research work, an efficient framework for monitoring the health of an individual and preventing the spread of airborne disease focused on coronavirus and its variant using machine learning approaches over spatio-temporal manner is proposed.

3.4 Deep Learning based Methods

Deep learning is an emerging technique that has been extensively utilized for prediction. The current presented work consisting various deep-learning neural network algorithms and machine learning models for diagosing the patients, suffering from airborne disease. DL can be supervised, semi-supervised, or unsupervised as ML. DL techniques include Convolutional Neural Network (CNN), Deep Belief Network (DBN), Deep NN, and Recurrent Neural Network (RNN). Some deep learning neural networks are ahead:

1. Recurrent Neural Network (RNN) - RNN is a type of artificial neural network designed to effectively model sequential data by capturing temporal dependencies. Unlike feedforward neural networks, which process inputs in a single pass without any internal memory, RNNs possess internal memory that allows them to remember past information and incorporate it into the current computation. Figure 3.1 represents the conceptual view of recurrent neural network.

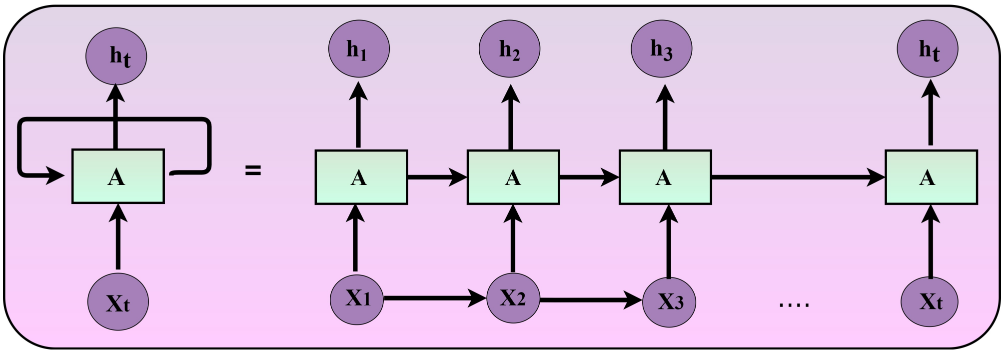


FIGURE 3.1: Architecture of Recurrent Neural Network

2. Convolutional Neural Network (CNN) - CNN is a type of artificial neural network designed specifically for analyzing visual data. It has proven to be extremely effective in tasks such as image recognition and classification, object detection, and even in more complex tasks like image segmentation

and facial recognition. Mainly three significant layers are included by the CNN namely; (i) Convolution layers, (ii) Pooling layers (PLs), and (iii)Fully Connected and activation function layer

3.5 Methodology

Research methodology is a way to find the solution of research problem systematically. It is a procedure used by researchers to present their work, data extraction, and explanation of the phenomenon. The researcher has taken the methodology process for solving the problem. Figure 4.1 described the step-by-step methodology process to achieve the objectives.

In this thesis, the research work is divided into two parts; the first part described the framework which is utilized for the identification of Coronavirus.

The second part of presented work described the another research work that helps to diagnose and identify the another airborne disease Omicron which is variant of coronavirus.

3.6 Data Preprocessing

Data preprocessing is a crucial step in machine learning (ML) pipelines. It involves transforming raw data into a format suitable for training ML models. The following various techniques used in data preprocessing are discussed ahead;

1. **Data Cleaning** - Missing data can adversely affect the performance of machine learning model. [127]. The strategies that are used for handling missing values include:

    - **Imputation -** Filling missing values using statistical methods like mean, median, or mode.

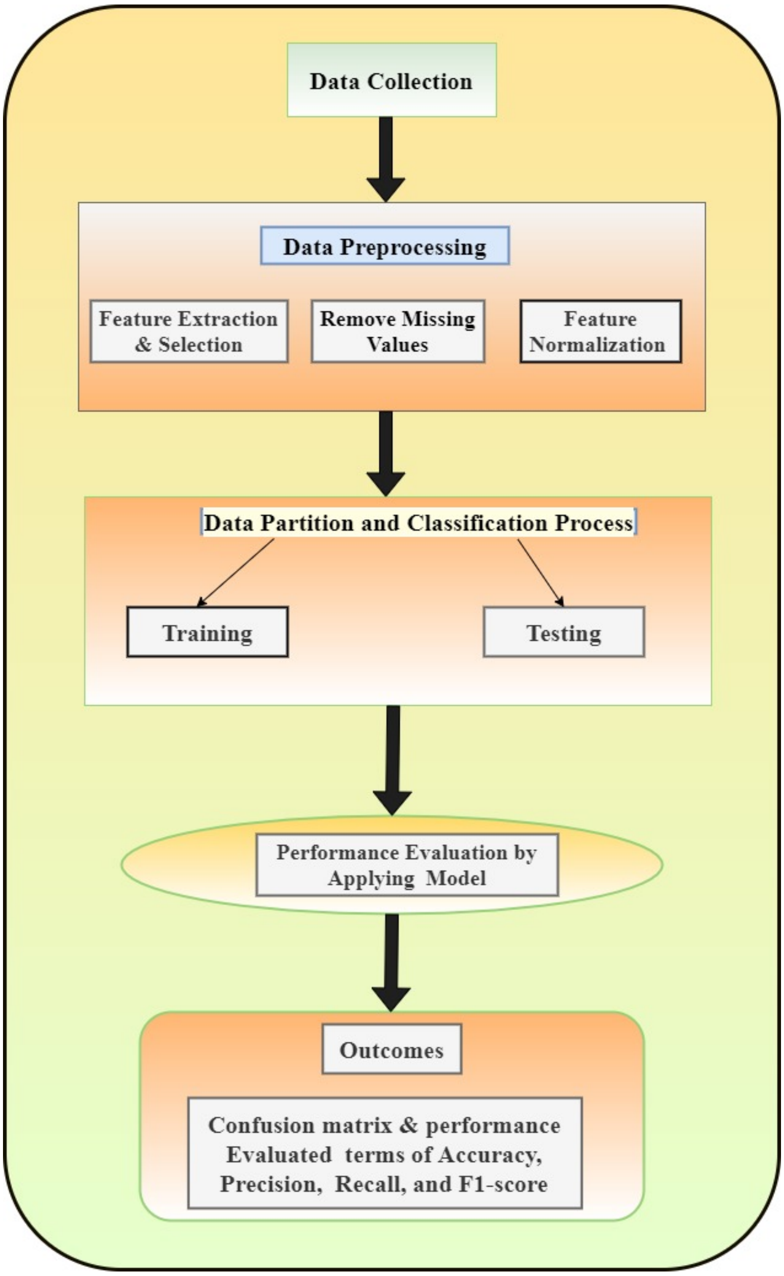


FIGURE 3.2: Workflow of the Research Methodology

- **Removing Outliers -** Detect and handle outliers in the data. Outliers can be treated by trimming, transforming, or removing them based on statistical methods that significantly deviate from the rest of the dataset.

- **Noise Removal -** Simply removing rows or columns with missing values.

2. **Data Transformation -** Transforming data to make it more suitable for modeling. It includes the following techniques:

   - **Feature Scaling (Normalization and standardization)-** Scale the features to a similar range to avoid dominance by certain features during model training. Common scaling techniques include standardization (subtracting mean and dividing by standard deviation) scaling numeric features to have zero mean and normalization (scaling features to a range between 0 and 1).
   - **Feature Encoding -** Convert categorical variables into numerical representations that can be understood by machine learning algorithms. This can involve techniques such as one-hot encoding, label encoding or ordinal encoding.
     (i) **One-hot Encoding -** Converting categorical variables into binary vectors.
     (ii) **Label Encoding-** Encoding categorical variables as integers.
   - **Feature Engineering -** Creating new features from existing ones or transforming existing features to better represent the underlying patterns in the data. This can involve operations such as polynomial features like square root to stabilize variance and logarithmic transformations (transforming skewed data distributions to be more symmetric).

3. **Feature Selection -** Selecting the most relevant features can improve model performance and reduce overfitting. Techniques for feature selection include uni-variate feature selection, recursive feature elimination, and feature importance ranking based on models like decision tree or random forest. The common methods are include:

- **Selection -** Identify and select the most relevant features that contribute the most to the prediction task. This can be done using techniques such as correlation analysis, feature importance ranking, or model-based selection methods.

4. **Dimension Reduction -** Reduce the number of features while preserving the most important information to improve model performance and reduce computational complexity. Techniques like Principal Component Analysis (PCA) is commonly used for dimensionality reduction.

5. **Data Splitting -** Split the dataset into training, validation, and testing sets to evaluate model performance.
   (i) Training Set - It is used to train the model.
   (ii)Validation Set - It is used to tune hyperparameters and assess model performance during training.
   (iii) Test Set - It is used to evaluate the final model performance.
   Common division in the ratio of 70-30, 80-20, or 60-20-20 for training, validation, and test sets, respectively.

6. **Handling Imbalanced Data -** In classification tasks, imbalanced class distributions can lead to biased models. It deals with datasets where one class is significantly more prevalent than others. It include the following techniques:

   - **Resampling -** Oversampling the minority class or under sampling the majority class to balance the dataset.
   - **Synthetic Data Generation -** Generating synthetic samples for the minority class using algorithms specifically designed for imbalanced data like SMOTE (Synthetic Minority Over-sampling Technique).

7. **Normalization -** Normalizing the data ensures that different features have similar statistical properties. This can involve techniques like Z-score normalization or min-max scaling.

8. **Augmentation -** Augmentation can overcome and to prevent the problem of over-fitting that enhance the accuracy of the proposed model. The data augmentation technique is applied to increase the insufficient data in training. This technique includes the flipping (horizontal flipping), scaling, shearing, and zooming. It is used to generate additional training examples by applying random transformations or perturbations to the existing data. This helps increase the diversity of the training dataset and improve model generalization.

Data preprocessing is essential for ensuring that the data is in a suitable format for improving model performance, and avoiding issues such as overfitting.

3.7 Machine learning and deep learning Approaches

This section described a brief overview of some ML and DL approaches which has described ahead:

Artificial Neural Networks (ANN)- ANNs is utilized in various tasks such as pattern recognition, classification, and regression. It processes information through weighted connections and activation functions. Input data is fed into the network, propagated through hidden layers, and eventually produces an output.

Support Vector Machine (SVM)- It is a supervised non-stochastic binary linear classification ML technique that is utilized to categorize the input data. It solves the regression problems. It helps to categorize the data by forming a line known as a hyperplane.

Random Forest Regression (RFRR)- RFRR is the best approach for the decision-making algorithms that create various decisions from an input dataset.

Such an algorithm divides the datasets into various sub-parts before creating further sub-trees.

K-nearest neighbors (KNN)- Such algorithm is used for both classification and regression tasks. To classify a new data point, the algorithm calculates distances between it and all other points in the training dataset often using Euclidean distance.

Support Vector Regression (SVR)- It is a regression algorithm based on support vector machine. It aims to find a function that predicts continuous output values for given input values. Feature scaling is crucial in support vector regression to enhance model performance by normalizing the data. Support vector regression involves support vectors, margin, and kernel functions for effective regression.

3.8 Conclusion

This chapter described the problem formulation and research methodology used in this thesis for identifying and diagnosing the airborne disease.

Chapter 4

Research Methodology for Coronavirus

4.1 Introduction

This chapter described the first research work that is related for diagnosing the coronavirus. In 2020, the pandemic of coronavirus has spread all over the World. Coronavirus is an airborne disease that in results ill effects on humans are recognized as public health concern globally. The main objective of this research work was to diagnose and prevent the spread of such disease timely. Spatio-temporal based deep learning model is used for detecting such disease so that the prevention measures could be taken on time. Deep learning is an emerging technique that has an extensive approach to prediction. In this thesis, the presented framework has used for the identification of patients that has been suffering from coronavirus so that necessary action can be taken promptly.

4.2 Research Methodology

This section described a deep learning based framework that has been utilized for the identification of Coronavirus. The step-by-step methodology process is described in Figure 4.1 for solving the problem.

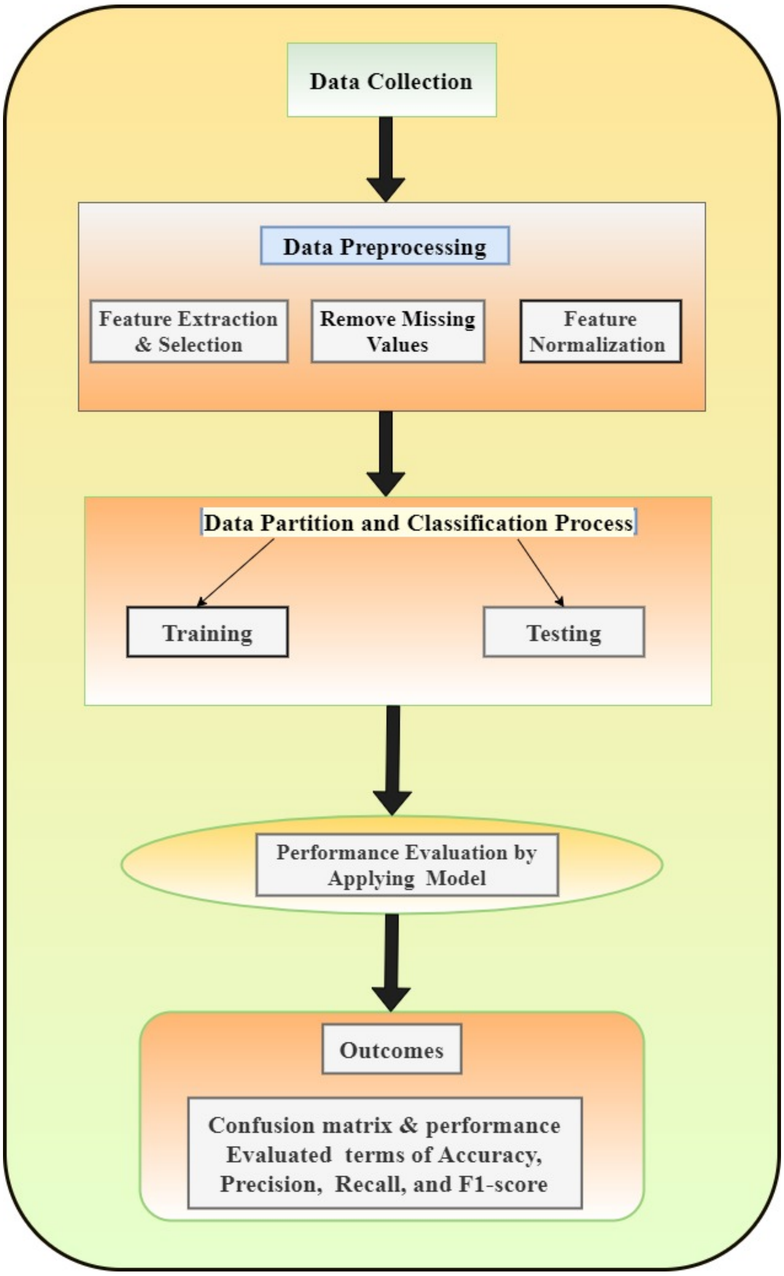


FIGURE 4.1: Workflow of the Research Methodology

4.3 Research methodology for the identification of coronavirus

The proposed methodology for the identification of coronavirus is summarized in Figure 4.2. In the first phase, data are collected from the publicly available GitHub Repository. Then, data are pre-processed to remove redundancy and augmented approach is used to remove the data imbalance and solve the problem of overfitting. The data extraction method is used for extraction features and used by the learned weights [128]. Next, datasets are divided into training and testing. Finally, the suggested model has tested with the dataset, to evaluate the model performance. The various statistical parameters has been evaluated in terms of accuracy, specificity, recall, and f-measure to explore the efficiency of the proposed model.

It includes the following steps:

1. Chest X-ray images from the datasets will be processed,
2. Then data augmentation will be apply.
3. Using pre-trained six models ResNetV2, GoogleNet, VGG2, ResNet152, AlexNet, and DenseNet512 were applied as transfer learning to detect coronavirus.
4. Next, Data is splitted into training and testing in the ratio 80:10:10.
5. After splitting of dataset, the deep learning based SQNet classifier is applied to diagnose the coronavirus class.
6. In the next and last phase the performance of implemented framework evaluated in terms of Accuracy, Recall, Precision, and F-measure Values.

Image Pre-processing - As the image input size varies, one of the significant phases in data pre-processing was to resize the X-ray images. Pre-processing used;

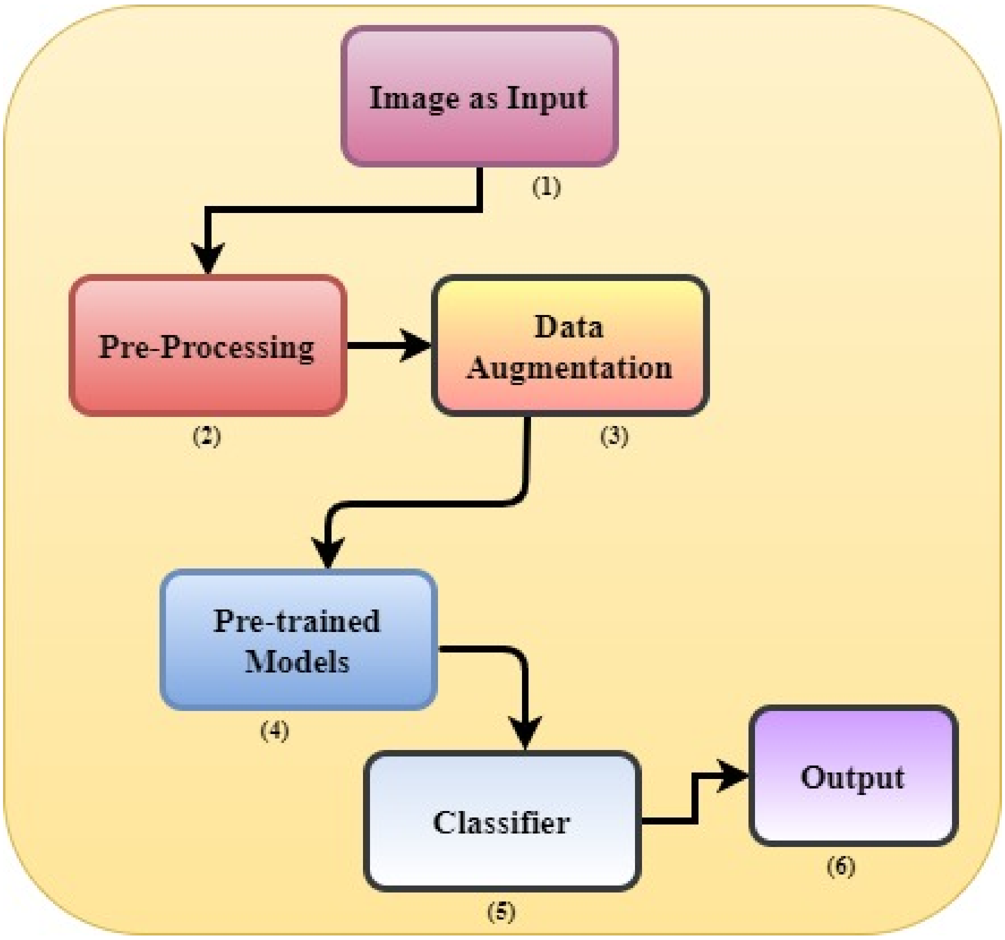


FIGURE 4.2: Step-by-Step Representation for the Identification of Coronavirus

1. Reshaping the image i.e. horizontal flip=true, scaling factor (1/255), shear range = 1/5, zoom range=1/5. The input of the whole network has 227x227 dimensions and consists of RGB channels. All images are resized to 227x227x3 dimensions in COVID-datasets.

2. Next, apply the image to an input of the pre-trained models.

3. Further, fetch the output of the last convolution layer of the given model.

4. Then flatten dimensions with reducing n dimensions to n-1.

5. Finally apply the SQNet model for the classification.

In the pre-processing step, it is required to label the data since the learning technique of convolution neural network fits into administered learning in machine learning. The pre-processing technique is implemented to increase the performance of the proposed system by speeding up the training time.

Image Augmentation - It could be possible that every problem does not have enough data, especially in medical-related issues. Sometimes collecting medical-related data may be time consuming and expensive. To solve these kinds of difficulties, augmentation can be applied. Augmentation can overcome and to prevent the problem of over-fitting that enhance the accuracy of the proposed model. The data augmentation technique is applied to increase the insufficient data in training. This technique includes the flipping (horizontal flipping), scaling, shearing, and zooming.

In this work, the research work comprises into different phases namely, (i) Data Collection, (ii) Data pre-processing and apply augmentation, (iii) Data division (Training and Testing Phase) and classification process, and (iv) finally the evaluation phase of building.

4.4 Step-by-Step Approach for the Identification of Coronavirus

This section presents a brief overview of the research framework which consists of the layered architecture based on the deep conventional neural network. In this thesis, the deep CNN based SQueezeNet model has been used for identifying the suspected cases of coronavirus disease. Figure 4.3 depicts the conceptual view of CNN based model to diagnose the CorV disease. The CNN comprises 3 different layers: (i) Convolution layers, (ii) Pooling layers (PLs), and (iii) Fully Connected and activation function layer. A brief description of each layer is described ahead:

4.4.1 Convolutional Layers

The core building blocks of CNNs are convolutional layers. As CNN is a multi-layer, each layer is responsible for extracting image features which are passed to the next layer. These layers apply a series of filters or kernels to the input image. With these filters, such layers are used to extract low and high-level pattern characteristics. Each filter in the CNN layer is utilized for the feature extraction process. Convolution involves sliding these filters over the input image, performing element-wise multiplication and summation operations to produce feature maps. The learning parameters of each layer consist 3×3 shaped matrix kernel which is used to transform the input matrix. When an image is passed through a filter, the value from filters creates a feature map. The stride parameter is used for adjusting the number of steps for shifting across the input matrix [129]. The output of the CNN layer is calculated by using the following formula:

$$Z_j^k = f\left(\sum_{m=1}^{N} w_j^{k-1} * y_m^{k-1} + b_j^k\right) \tag{4.1}$$

where, Z_j^k is feature map as output in a layer.
w_j^{k-1} indicates the jth filter and y_m^{k-1} represents the feature map in k-1 layer respectively.
b_j^k indicate the bias of the jth feature map at k^{th} layer.
N represents the total number of features at $(k-1)^{th}$ layer.
(*) indicate the vector convolution process.

4.4.2 Pooling Layers

After convolution, pooling layers are often used to diminish the spatial size and number of feature maps while retaining the most important information. The Max pooling (MXP) and Global average (GA) pooling methods are used to reduce the dimensionality and complexity in this work. MXP is basically used inbetween two

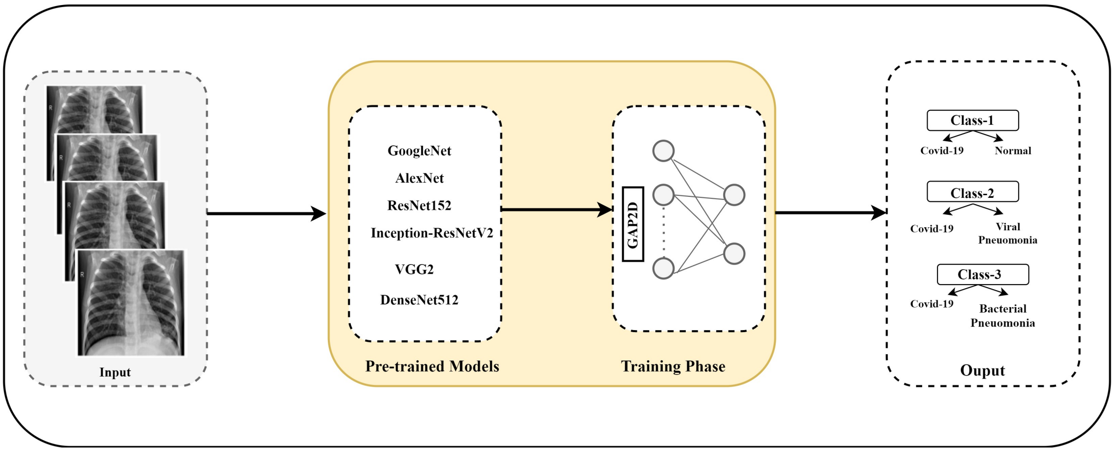


Figure 4.3: Conceptual view of the proposed Model

convolution layers which separate the input. GA pooling selects the highest value as output in feature map by utilizing the matrix given by the neurons as output. The GA pooling determine the average to reduce data into a single dimension [130]. The flattering layer gathers data into a single vector and forward it to the fully connected layer.

4.4.3 Fully Connected Layers(FCLs)

One or more fully connected layers may be added at the end of the network. These layers act as classifiers, taking the high-level features extracted by the convolutional layers and making predictions based on them. After each convolutional operation, an activation function ReLU (Rectified Linear Unit) is applied element-wise to introduce non-linearity into the network. Moreover, the softmax function is used as the output to predict the output image into a probability distribution over the different classes. In FCL, all neurons are fully connected to all activation functions in the previous layer. Both ReLU and Softmax function are mathematically represented as:

$$ReLU(z) = \begin{cases} 0, & \text{if } z < 0. \\ z, & \text{if } z \geq 0. \end{cases} \tag{4.2}$$

$$Softmax(z_i) = \frac{e^{x_i}}{\sum_{j=1}^{c} e^{x_j}} \tag{4.3}$$

where x_i represents the inputs and c indicates the number of classes.

4.5 Transfer Learning

Transfer learning (TL) is a machine learning technique where a model trained on one task is transformed for another related task. Rather than starting the learning process from scratch, TL used the gained knowledge to solve one problem and applies it to a different related problem. There are various ways to perform transfer learning such as feature extraction and fine-tuning.

Nowadays, TL approach is more useful in the field of medical applications. The primary benefit of using the TL strategy is that it involves fewer computations and it allows data training with a smaller number of datasets. TL is the process of creating a new model using calculated network layer weights by applying knowledge from an existing pre-trained model [131]. Therefore, TL is used to train newer models in the destination domain using the knowledge, weights, and features, acquired from the source domain [132].

In this work, TL is often used with CNN in the way that all layers are kept except the last one, which is trained for the specific problem. Deep learning models often need a lot of data. Labeling this data by experts is both costly and time-consuming. The biggest advantage of using the TL method is that it allows the training of data with fewer datasets and requires less calculation costs. With the transfer learning method, which is widely used in the field of deep learning, the information gained by the pre-trained model on a large dataset is transferred to the model to be trained. This eliminates the requirement of having a large dataset and also reduces the training time.

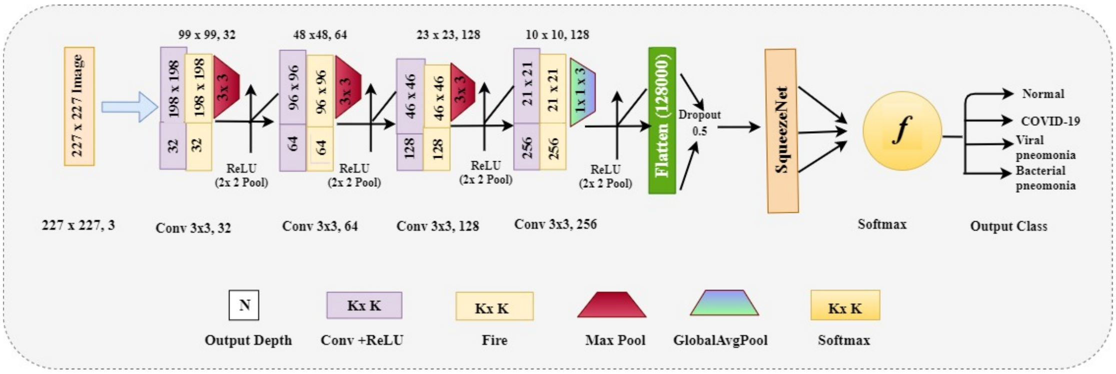


FIGURE 4.4: Architecture of the Proposed Work

In this work, TL has been used to achieve the information provided by the pre-trained models, transferred on a large dataset to be trained to the model. Using pre-trained six models ResNetV2, GoogleNet, VGG2, ResNet152, AlexNet,and DenseNet512 were applied as transfer learning to detect coronavirus.

4.6 Proposed Framework (SQNet) For the Identification of Coronavirus

In this thesis, the deep CNN based SQueezeNet model has been used for predicting the CorV disease [133]. In the presented work, a deep CNN-based SQNet model is utilized to predict the CorV disease. SQNet framework comprises of convolution layers, MAX pooling, GA pooling layer, fire levels, and softmax output layer. Figure 4.4 represents the architecture of the proposed approach. The Stride (s), filter size (K x K), and length(L)is used for the feature map respectively. The input image consists of RGB channels with the dimension size 227 x 227. To make the input images more generalize, convolution is used. CNN's first layer convolves the input's small area and weights using 3×3 kernels. SQNet uses the fire layers which comprises of (i)squeeze and (ii) expansion stages between the convolution layers. In the squeeze phase, the size 1×1 filter and in the expansion phase, the sizes 1×1 and 3×3 filters are used. Figure 4.5 represents the squeeze and expansion phase of the SQNet architecture. The total number of convolution channels

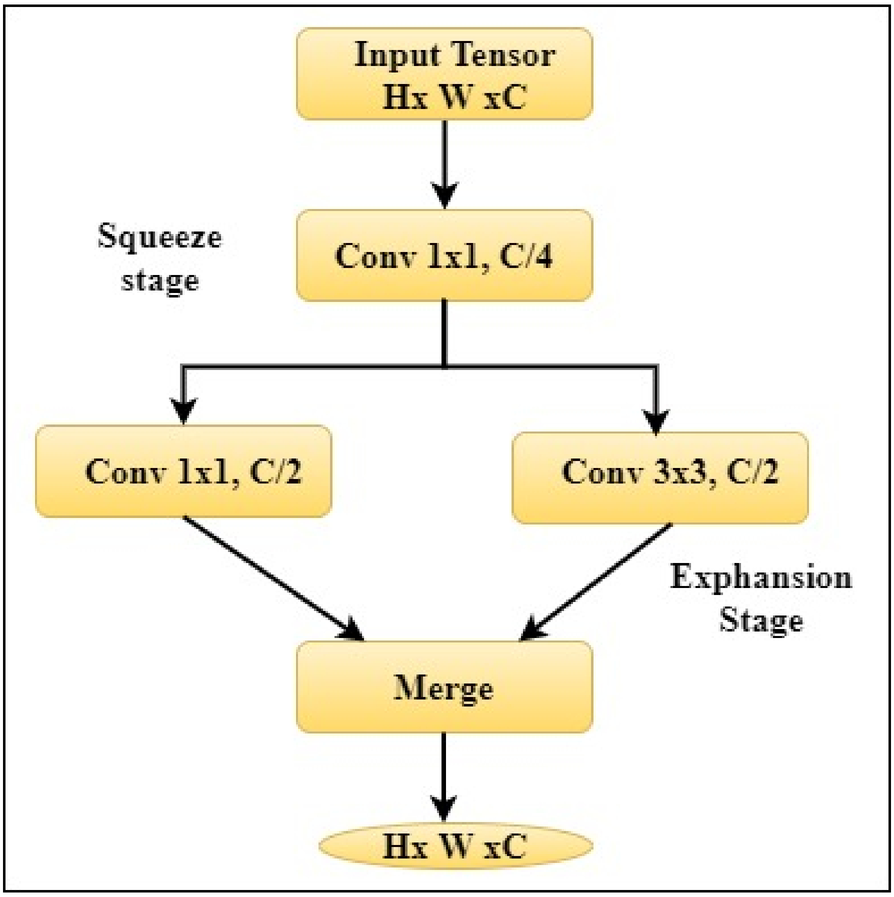


FIGURE 4.5: Squeeze and Expansion Phase of Fire Layer

equals to the number of input tensors(Hx W x C). Both the squeezing and expanding stages consist the ReLU. The squeezing process reduces the depth, whereas the expanded phase increases it for retaining the same feature size. Finally, the expanded outputs are stacked in the dimension of the input tensor by utilizing the merging operation.

The proposed model follows numerous steps for predicting the CorV disease. After the initial step, the data goes through the expansions, and its depth is increased to the output tensor. Finally, the resultant f(y) of the squeezing process with the kernel(W), feature maps(FM), and C specify channels of different tensors respectively, which can be mathematically represented as:

$$\text{f(y)} = \sum_{fm=1}^{FM} \sum_{c=1}^{C} W_c^f * x_c^{fm} \qquad (4.4)$$

Where f(y) is an output $\in R^N$.

Let X_i be input X of layer i with a size of $(W_i, H_i, C_i) \in R^N$ where W represent to the height, H defines the height, and C indicate the channel respectively. Moreover, Table 4.1 described the detail description of proposed model with the dropping rate of 0.5. The feature map is diminish into a single value in the network by Max pooling and Global average pooling which helps to converts the feature maps into a single value. In the last layer, the Softmax activation function produces multi-class probability distributions. Multiclass classifiers are used to classify images into a collection of categories. These are also referred to as output layers in neural networks. For the classification process, the classifier needs

Table 4.1: Detailed Architecture of the Presented Model

Layer	**Layer_Type**	**Activation**	**Stride**	**Output shape**
Conv_2d	Conv+ReLU	3x3	2	198x198x32
pool1	MaxPooling	3x3	2	198x198x32
Fire	Conv+ReLU(expand+merge)	1x1	1	128x128x32
	(Conv+ReLU(expand+merge))/2	3x3	1	128x128x64
pool2	MaxPooling	3x3	2	128x128x64
Fire	Conv+ReLU(expand+merge)	1x1	1	56x56x32
	(Conv+ReLU(expand+merge))/2	3x3	1	56x56xx64
pool3	MaxPooling	3x3	2	99x99x128
Fire	Conv+ReLU(expand+merge)	1x1	1	99x99x128
	(Conv+ReLU(expand+merge))/2	3x3	1	28x28xx128
pool4	MaxPooling	3x3	2	56x56x64
Fire	Conv+ReLU(expand+merge)	1x1	1	28x28xx256
	(Conv+ReLU(expand+merge))/2	3x3	1	14x14xx256
Conv_2d	Conv+ReLU	3x3	1	14x14x128
pool	Global Average Pool	-	-	1x1x3
Output	SoftMax	-	-	1x1x3

unique features. Consequently, the feature extractor output is converted into a 1-dimensional feature vector for the classifiers. Flattening is the process of creating a single, lengthy feature vector from the convolution operation that the SQNet layer uses for its last classification step.

4.7 Experimental Design

The current section describes the experimental setup for the proposed framework. For result generation, the suggested model was implemented in a Matrix laboratory (MATLAB) toolkit. The experiment for coronavirus was conducted running in MatLAB environment on computer system with the speed of 3.3@GHz, Intel Xeon E5, Quadro M4000 8GB GDDR6 graphic card , and memory capacity of 512 GB. The experimental results of proposed system are compared with state-of-the-art(SOA) models.

4.8 Dataset Collection and Classification

The present chapter provides detailed information about data collection and methods used in this research work. This chapter also explains the sources from where data has been collected. The following methods are explored and analyzed to complete the task which are given in the further sections.

4.8.1 Data Collection Approaches

Data collection is the process of collecting data, quantifying, and evaluating it through the use of a set of verified standard methodologies. This task is accomplished using extraction and collection of the necessary data for analysis as accurate data is the primary substantial part of the analysis. The main objective of data collection is to collect comprehensive and authentic data and analyze it to make key decisions. After the collection of data, it undergoes a data cleansing process and processing process to ensure that it is genuinely useful. The data-gathering technique is categorized into two types (i) Primary data collection (ii) Secondary data collection methods. Each method is discussed as follows:

1. **Primary Data Collection Method -** A primary data collection is the first source where data is obtained directly from a specific experiments, surveys, observations, or clinical reports. The primary data collection is further defined as:

    - Quantitative Data Collection Method: It is based on mathematical calculations and collected in a variety of forms such as questionnaires, correlation, and regression analysis techniques. This method is less expensive and more efficient than qualitative data collection technique as it cannot be easily altered by a human being.

    - Qualitative Data Collection Method: There are no mathematical computations at all. Non-quantifiable components are intrinsically tied to this methodology. Case studies, observations, questionnaires, and interviews are a few types of qualitative data collecting techniques. The various techniques for collecting this type of data are: data sources on Internet, Interview method, and observation method.

2. **Secondary Data Collection Method -** The term ”secondary data” means the data is already used in another research. In this case, it signifies that the data is already available from secondary sources such as newspapers, books, and journals. It might be previously published data or previously unreleased data. Published data can be from government publications.

In the research work for coronavirus, the publicly available dataset collected from open source repository GitHub is utilized.

4.8.2 Data Sources for Coronavirus

The details of the datasets sources have been described in this current section. The three different datasets which comprises X-ray images of infected, non-infected, bacterial and viral pneumonia class. These datasets consists chest X-ray images mostly from patients suffering from SARS, CorV, MERS, and Pneumonia collected from the GitHub publicly available website shared by Dr Cohen [134]. In addition, 165 images are selected from the "covid-chestxray-dataset" Dataset [1]. The Experiment has been performed on three datasets namely (a) Cov-Dataset1, (b) Cov-Dataset2 (iii) and, Cov-Dataset3.

4.8.3 Dataset Classification

In research work, to diagnose and identify the coronavirus these datasets
(a) Cov-Dataset1,
(b) Cov-Dataset2
(iii) and, Cov-Dataset3 are utilized in the experimental study.
The distribution of images per dataset is described in Table 4.2. The collected

Table 4.2: Number of images in each dataset

Datasets	**Healthy**	**Bacterial**	**COVID-19**	**Viral**
Cov-Dataset1	2860	–	126	–
Cov-Dataset2	–	2228	126	–
Cov-Dataset2	–	–	126	3517

---

[1] source:https://github.com/ieee8023/covid-chestxray-dataset/blob/master/metadata.csv

datasets consist of a total of 8731 images with 7662 patients of disease-infected cases, 2860 Normal, 2228 Bacterial pneumonia, 3517 Viral, and 126 CorV disease cases.

4.9 Research Outcomes

As described above, the research framework for coronavirus comprises the layered architecture. The datasets are collected the from the UCI that were obtained from the GitHub repository. The datasets for coronavirus consist of X-ray images of MERS, SARS, Viral, and Bacterial Pneumonia patients. The CNN-based SQNet model applied to categorise the data into possible categories Normal, COVID-19, Bacterial pneumonia, or Viral pneumonia. A total of 8731 images consisting of three benchmark datasets were analyzed for the prediction of CorV patients. Moreover, the omicron datasets consits of 9525 disease-infected cases, 2851 healthy, and 2500 other cases.

Furthermore, the effectiveness and accuracy of the proposed system are discussed in this chapter. In addition, the prediction efficacy of the implemented framework is explored with many statistical metrics like accuracy, precision, f-measure, and recall. To enhance the prediction performance, the presented work was evaluate with various statistical errors and to evaluate the efficiency of implemented framework has compared with other prediction models.

4.9.1 Statistical quantitative Parameters

The current subsection shown the quantitative-wise performance of the proposed approach in terms of accuracy, truthfulness (TR), faultless (FLT), specificity, f-measure. Additionally, the statistical measure Matthew correlation coefficient (MCR) are also evaluated for the performance of the proposed model. These metrics perform the following functions:

1. **Accuracy** - It is used to determine the overall classification performance of the proposed model. It shows the ratio of the correct predictions through classification from a total dataset.

   Accuracy = Number of correct Prediction/ Total number of Predictions

2. **Precision** - It determine the frequency of positive class occurring in the model. In this thesis, it specifies the rate of correctly predicted the actually disease infected.

3. **F-measure** - It calculate the harmonic of recall and precision. It is appropriate for calculating imbalanced data performance.

   - TR - specify the actual rate of classification of the dataset.
   - FLT - specify the true rate of negative data.

4. **MCR** - It specify the quality-wise classification performance.

5. **Negative predictive value (NPV)** - defines the probability of how truly classified the result that not affected with the disease actually. NPV increases shows that there will be more true negatives for each false negative.

4.9.2 Confusion Matrix

A confusion matrix is a tool to evaluate the classifiers. It shows the relationship between true positive and true negative. In confusion matrix, the four parameters can be expressed as:

1. **True Positive (TP)** - represents the number of truly identify instances with the disease. Those instances where predicted class and actual class both are positive called TP.

2. **True Negative (TN)** - identify the total number of truly instances with another disease. In other words, those instances where predicted class and actual class are both are negative named as TN.

3. **False Positive (FP)** - defines the number of wrongly instances with disease. In simple words, those instances where a predicted class is positive, but instances are negative is called FP.

4. **False Negative (FN)** - defines the number of incorrect instances other than disease. In other words, those instances where a predicted class is negative, but the instances are positive, is known as FN.

Figure 4.6 represents the confusion matrix in which each row state the actual class and each column represents the predicted value of each class.

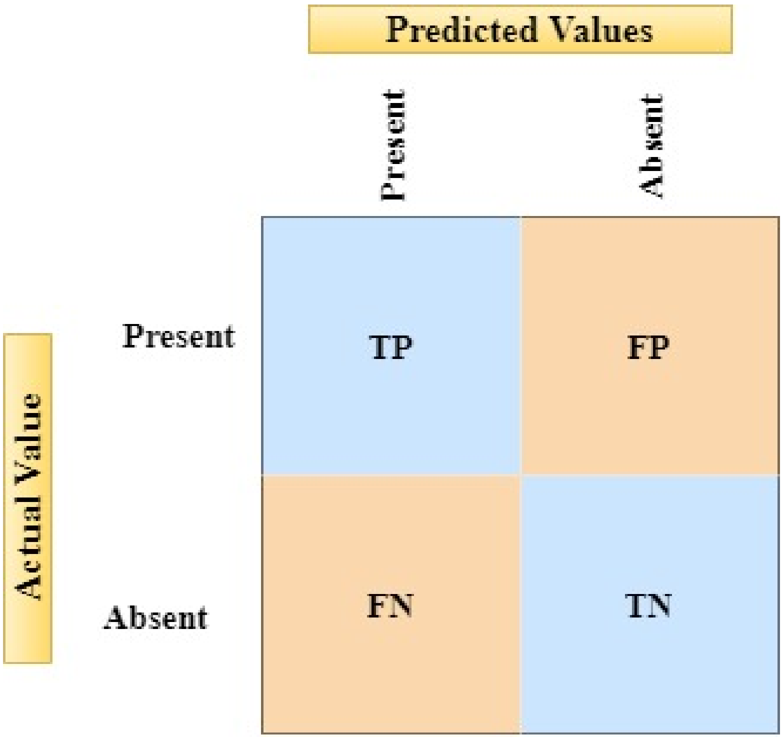


FIGURE 4.6: Confusion Matrix

According to the confusion matrix, the performance evaluation parameters are calculated as:

$$ACR = \frac{N_{TRP}+N_{TRN}}{N_{TRP}+N_{TRP}+N_{FLP}+N_{FLN}}$$

$TR/PPV = \frac{N_{TRP}}{N_{TRP}+N_{FLP}}$

$FLT = \frac{N_{TRP}}{N_{TRP}+N_{FLN}}$

$SPFC = \frac{N_{TRN}}{N_{TRN}+N_{FLP}}$

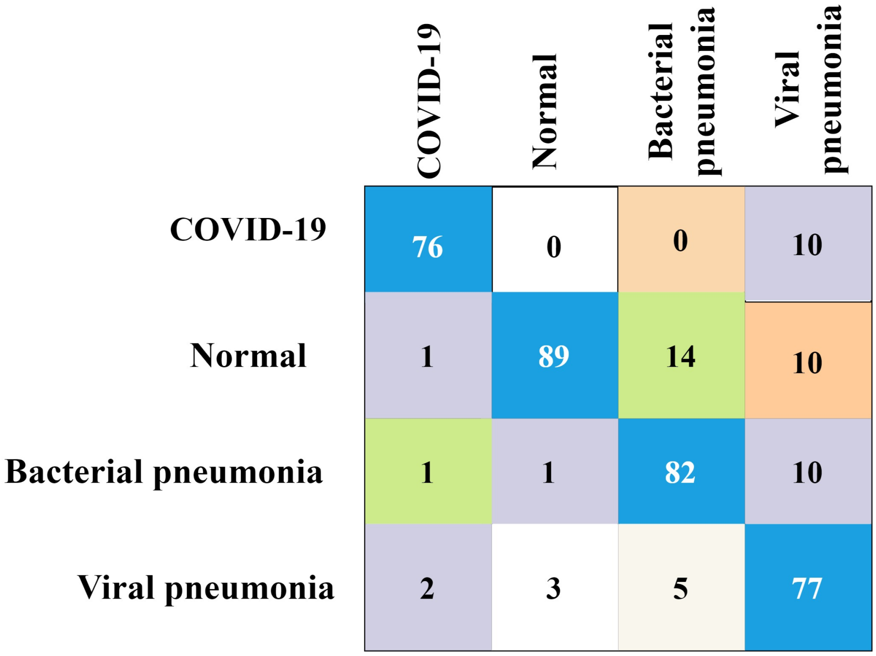
$MCR = \frac{(N_{TRP}*N_{TRN})-(N_{FLP}*N_{FLN})}{\sqrt{(N_{TRP}+N_{FLP})*(N_{TRP}+N_{FLN})*(N_{TRN}+N_{FLP})*(N_{TRN}+N_{FLN})}}$

$F - measure = 2 * \frac{TR*FLT}{FLT+TR}$
NPV can be calculated by the formula:
$NPV = \frac{TN}{TN+FN}$ (1)

Here $N_{TRP}, N_{TRN}, N_{FLP}, N_{FLN}$ represented the number of truly classified, number of correctly define snegative cases, number of incorrectly defines positive cases, number of incorrectly defines negative cases with the disease respectively. By using such evaluation metrics, the classification method of the proposed model enhances the performance and effectiveness to determine the coronavirus AB disease. Figure 4.7 represents the confusion matrix of 4-class distribution of dataset Cov-dataset2
.

4.9.3 Performance Analysis

Validation is the process of ensuring that the results obtained from experiments meet our expected outcomes and appropriate. The difference between the predicted outcomes and actual results is called the training error.



	COVID-19	Normal	Bacterial pneumonia	Viral pneumonia
COVID-19	76	0	0	10
Normal	1	89	14	10
Bacterial pneumonia	1	1	82	10
Viral pneumonia	2	3	5	77

FIGURE 4.7: Confusion Matrix(4-class distribution)

4.9.3.1 Training, Testing, and Validation implementation For Coronavirus

This approach is most commonly used to validate the machine learning model. In this method, a dataset can be subdivided into (i) Training (ii) Testing and (iii)Validation. The most common approaches for training and testing are 70:30, 80:20, or 80:10:10 of a dataset. In this validation test, the main dataset is divided into K =5 folds for validation. In each iteration, each fold is selected as a test dataset to evaluate the predicted model.

In this research work, augmentation approach is applied to the raw dataset for fine-tuning as well as for end-to-end training. Figure 4.8 depicts the original images means before augmentation. In the training phase, the data augmentation approach with hyperparameters and their range has utilized as described in Table 4.3.

The original dataset (before augmentation) and augmented dataset (after augmentation) has been depicted in Table 4.4. After augmentation, all images data are

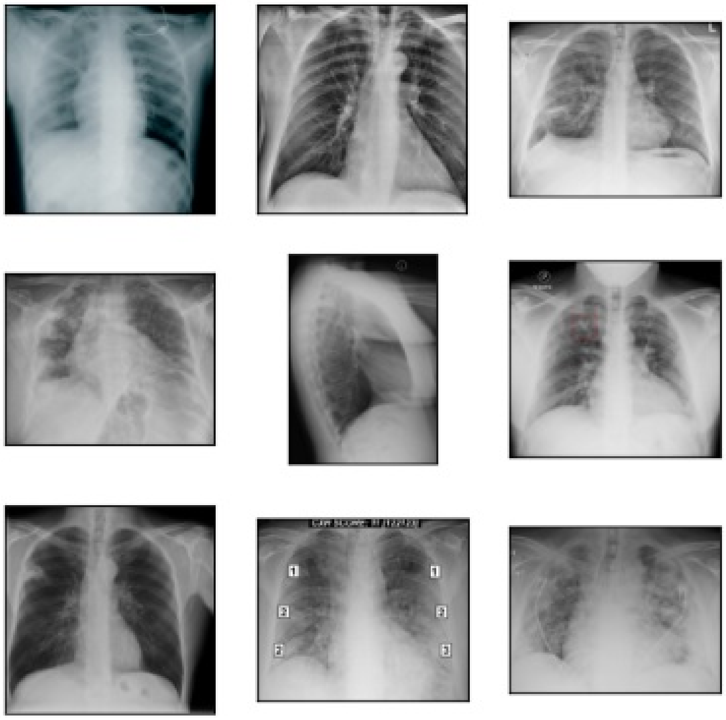


FIGURE 4.8: Images Before Augmentation

TABLE 4.3: Hyperparameters and Their Values

Hypeparameters	**Values**
Scaling	1/255
shearing	1/5
Zooming	1/5
Flipping/Rotation	Horizontal

scaled to 227 × 227 pixels for the obtained datasets. Figure 4.9 depicts the few images after augmentation by using hyperparameters. In the presented experimental, the k = 5-fold cross-validation is used to increase the validation performance as well as to achieve a robust result. Further, the obtained dataset is sub-divided into 80%, 10%, and 10% for training validation, and testing purposes respectively.

Training Accuracy - defines a how much data that is actually trained.

Validation Accuracy - is linked to training data that are actually trained.

Cross Entropy - described about loss function. It defined how finely the loss function has trained.

Table 4.4: Distribution of datasets Before and After Augmentation

Classes		pre augmentation			After augmentation	
	Train	**Validation**	**Test**	**Train**	**Validation**	**Test**
COVID-19	86	-	40	1580	114	113
Normal	1875	-	985	1580	114	113
Bacterial	1716	-	512	1580	114	113
Viral	2814	-	701	1580	114	113

For the purpose of evaluation performance, The training and validation datasets are used by the proposed model. To achieve the optimization result, the adaptive moment estimation (ADAM) optimizer was utilized during the training of the data. The proposed model learn rate 0.001 at the starting of training which was decreased after two epochs. The cross-entropy loss function was utilized during training process.

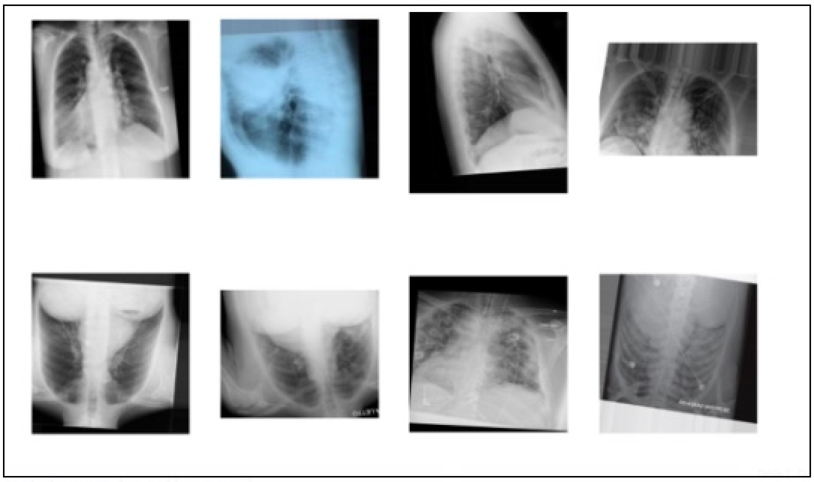


Figure 4.9: After Augmentation

Table 4.5: Results of Statistical parameters (Four-class classification(in %)

Classes/Metrics	ACR	TR	FLT	SPFC	F-measure	MCR
COVID-19	96.71	96.2	94.13	99.6	95.12	62.7
Normal	98.6	93.4	98.52	97.4	96.71	58.09
Bacterial pneumonia	97.5	87.21	86.7	94.1	88.51	56.11
Viral pneumonia	97.7	86.1	83.2	94.3	88.3	56.2
Average	97.8	92.7	91	96	92.7	56.0

4.9.3.2 Result Analysis

To evaluate the classification performance of the proposed model, the important parameters metrics accuracy, truthfulness, faultless, specificity, f-measure, and statistical measure Matthew correlation coefficient has been computed by the utilization of confusion matrix. The average accuracy, specificity, f-measure, and MCR can be seen in Table 4.5.

4.9.3.3 Research Outcomes

1) As seen in Figure 4.10, the suggested model acquires a higher accuracy of 97.4%, ResNet has an accuracy of 91.1%, and GoogleNet has an accuracy of 90.9% for the obtained datasets. The average accuracy has been achieved by the suggested model consisting of a value of 97.4%.

2) In terms of specificity analysis, the suggested model has a higher value of 93.65% than ResNet (90.9%), and VGG2(91.2%). Consequently, the provided model has obtained a higher value than other models with an average specificity consisting of 96% which is significantly superior.

3) The evaluation of the effectiveness of the suggested model also includes the analysis of the f-measure. As it can be seen in Figure 4.12, the provided model

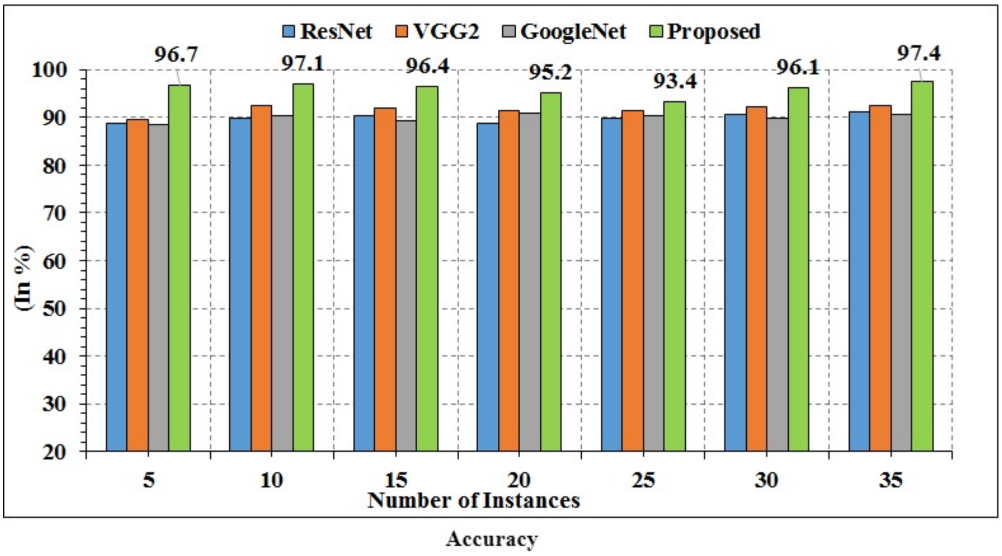


FIGURE 4.10: Accuracy

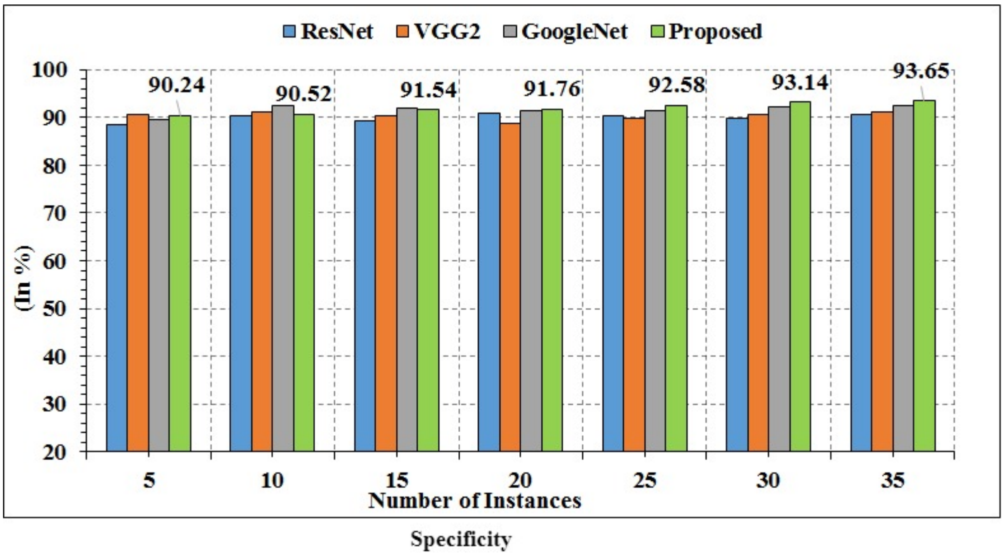


FIGURE 4.11: Specificity

obtains a high value of 95.12% as compared to other approaches. Based on the evaluation analysis from the datasets, it is demonstrated that the suggested model is more effective and efficient.

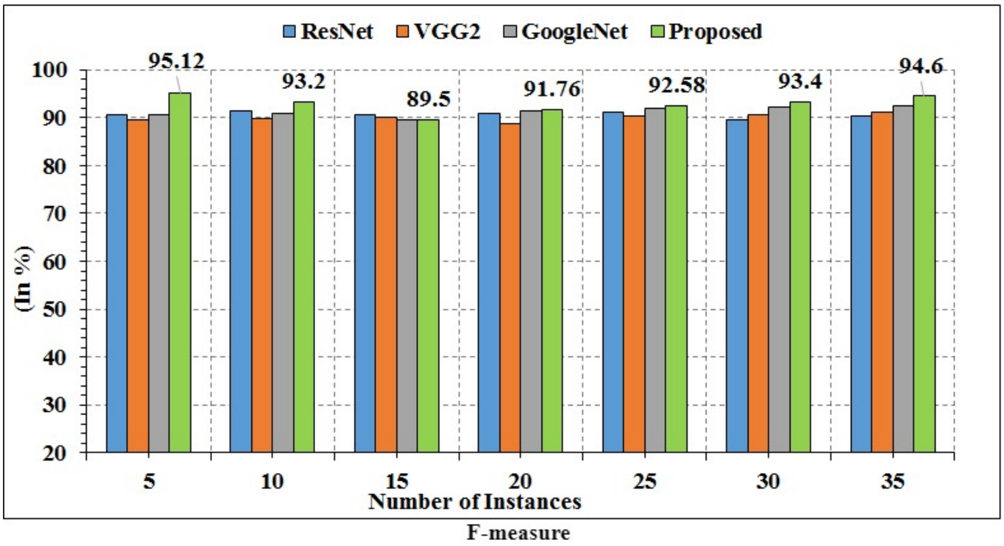


FIGURE 4.12: F-measure

4.9.4 Prediction Efficiency

Prediction efficacy measures the precision of the proposed approach in the predictivity of disease.

As seen in Figure 4.13, the proposed model has obtained an average higher accuracy of 97.8% that represented the validation accuracy of the proposed model. A higher recall (FLT) value represented the lesser number of false negatives (FLN) that is an encouragement to achieve better outcomes. The implemented model

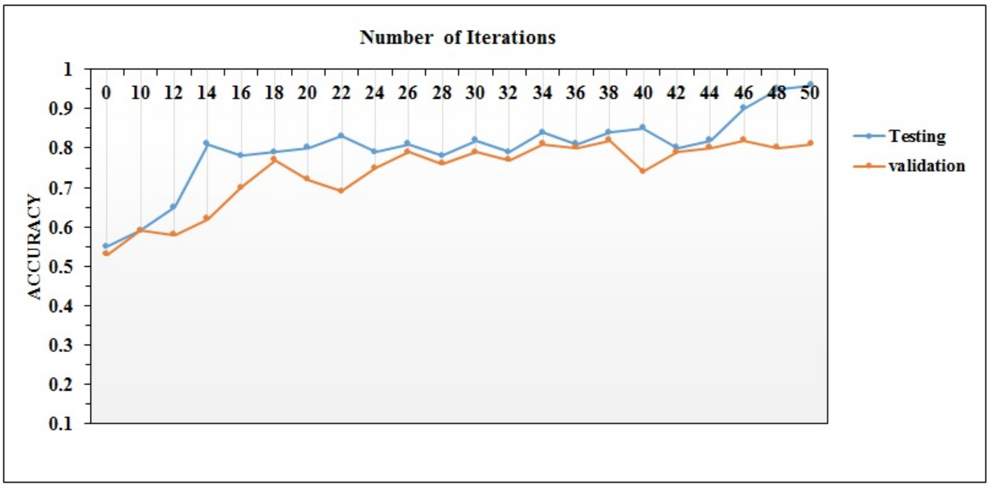


FIGURE 4.13: Proposed Model Accuracy Curve

Table 4.6: Number of cases used during Testing

3-Class	**Number of cases**
CorV-infected	45
Normal	76
Pneumonia cases	128

has been achieved average recall value 91.0% can be seen in Table 4.5 from the obtained datasets. Figure 4.14 has been shown the loss values for both training as well as testing with dropout rate 0.5.

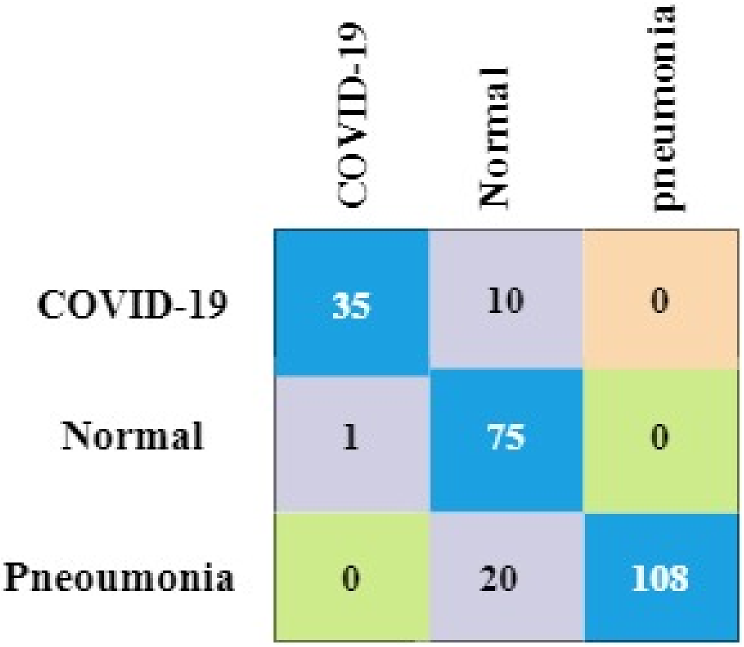
The model was tested according to the distribution cases depicted in Table 4.6 for possible classification after utilizing the fine-tuning approach. The batch set of size 36 was used in the training phase for 100 iterations. The mean subtracting approach has been used to normalize all the images during the training phase.

Figure 4.14: Model Loss Curve

It was noticed that the performance achieved for pneumonia class not much better, compared to other classes. So, pneumonia and viral class were combined into a single class. The average accuracy obtained by the predictive model 98.47% which has increased as it can be seen in Table 4.7. The average computational time

TABLE 4.7: Outcomes of Various Metrics for 3-class classification(in %)

Classes	**ACR**	**TR**	**FLT**	**SPFC**	**F-measure**	**MCR**
COVID-19	99.22	98.1	88.9	94.0	98.5	71.8
Normal	98.7	93.2	86.51	90.1	97.41	58.09
Pneumonia	97.52	91.2	93.4	91.3	89.51	56.11
Average	98.4	94.2	89.6	91.8	95.1	61.9



	COVID-19	Normal	pneumonia
COVID-19	35	10	0
Normal	1	75	0
Pneoumonia	0	20	108

FIGURE 4.15: Confusion matrix for 3-classes Classification

for the implemented model on CPU was 0.347s and for GPU it was 0.152s. The confusion matrix for 3-class categorization has been shown in Figure 4.15.

4.10 Objective Function Evaluation

In this thesis, the deep SQNet model includes a validation step in addition to optimisation during the training phase. The objective function of the optimisation process for the objective function has displayed in Figure 4.16. As it can be seen

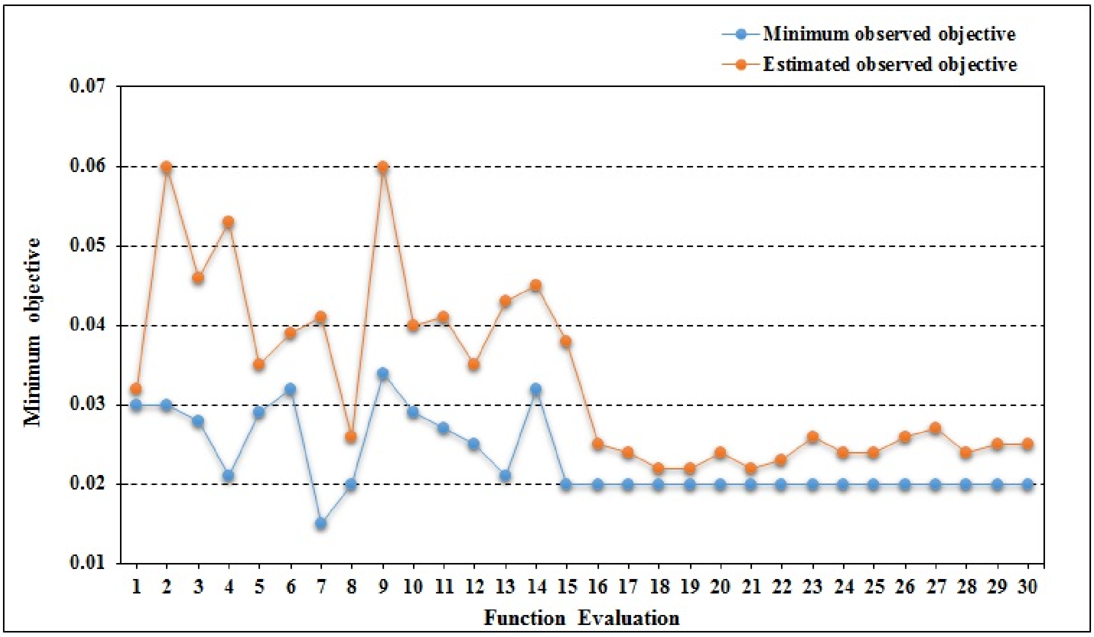


FIGURE 4.16: Objective Function Evaluation Analysis

in Figure 4.16 at the end of ninth iteration, the minimal observed aim to develop the best model has been achieved. The suggested model's performance boost is increased by about 20 times using the augmentation strategy.

4.11 Predication Efficiency Result with MCR

As a measure of the quality of binary classifications, a machine learning approach that employs the MCR has been utilized. It can be mathematically calculated using the following formula:

$$MCR = \frac{(N_{TRP} * N_{TRN}) - (N_{FLP} * N_{FLN})}{\sqrt{(N_{TRP} + N_{FLP}) * (N_{TRP} + N_{FLN}) * (N_{TRN} + N_{FLP}) * (N_{TRN} + N_{FLN})}} \quad (2)$$

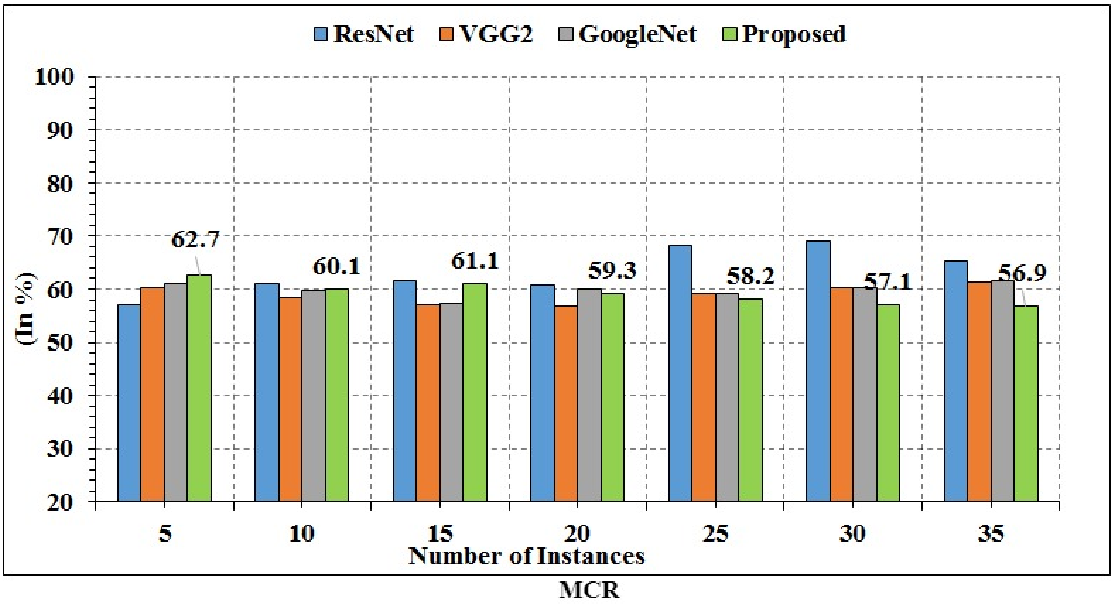


FIGURE 4.17: MCR Evaluation

Figure 4.17 has shown that the MCC for CorV class consist of a higher accuracy value of 71.8% from the obtained datasets for the prediction of coronavirus ABD.

4.12 Comparative Study With Other Research Works

The use of DL techniques in images classification procedures can aid in the early detection of illness. In terms of the deep neural network, CNN techniques outperform traditional diagnosis methods in terms of accuracy and efficiency. Deep SQNet is presented as a fast, reliable, and efficient COVID-19 diagnostic technique in this thesis. The presented approach classifies the images into Healthy, Pneumonia, Viral, and COVID-19 categories.

Li and Zhu [135] present a COVID-Xpert architecture based on DenseNet for identifying three-class chest X-ray images.In their work, the presented model had achieved an overall accuracy of 89.0% by using transfer learning. Wang and Wong [136] discussed the COVID-Net architecture for COVID-19 diagnosis, and the

Table 4.8: Comparative Studies

Metrics	**Class**	**Methods**	**Datasets**	**Augmentation/ Transfer learning**	**SPFC (%)**	**Recall (%)**	**Accuracy (%)**
Afshar et al.[83]	4	CapsuleNet	Public	Fine-tuned	96	90.1	95.7
Song et al.[98]	2	DRE-Net	Private	Transfer learning	-	93	98.5
Li and Zhu[135]	3	DenseNet	Repository	No	92	87	88.9
Ionnis et al.[84]	2	VGG-19	Public	Transfer learning	96.2	-	93.4
Ozturk et al.[93]	3	DarkNet	Public	No	90	92.1	87.1
Hariri et al.[137]	3	sgdm	Public	Transfer learning	-	-	98
Das et al. [138]	3	Xception	Public	Transfer learning	96.9	97.1	97.4
Wang et al.[136]	2	ResNet50	Public	Transfer learning	92.1	91	97.8
Narin et al.[91]	2	nCoVNet50	Public	Transfer learning	89.1	98.0	88.8
Presented System	4	SQNet	Repository	Both	96.0	91	98.4

primary model in their study is based on the customized CNN. The model architecture is improved via machine-driven design. Experiment results provide overall accuracy, COM, and COR scores of 92.2%, 88.7%, and 91.3%, respectively. The COVIDx dataset utilized is also shared and collected by the authors. Farooq and Hafeez [89] proposed a ResNet-based system for identification of COVID-19. The model's accuracy has remained constant at 0.962. Table 4.8 depicts that the proposed SQNet model has achieved a higher accuracy value of 98.4% when compared to other models.

4.13 Strength and Weakness Comparative Study With Other Recent Works

The presented SQNet model demonstrates competitive performance compared to recent studies, achieving an accuracy of 98.4% for multi-class classification. The presented approach has contributed to its robust performance for COVID-19 classification using chest X-ray images. However, like other complex models, it may require significant computational resources and further validation to ensure its effectiveness across diverse datasets. Table 4.9 has shown the strength and weakness comparison key points with other research works.

4.14 Discussion

In this work, a deep learning based CNN-based SQNet framework has implemented for effectively and accurately classifying the coronavirus cases from healthy, pneumonia cases by using obtained datasets. The implemented model has provided accurate diagnostics for binary as well as multi-class classification for the prediction of coronavirus. This chapter summarizes the researcher efforts to test and validate the prediction model for coronavirus. The datasets were obtained from Github repository. Moreover, the performance measure metrics have been evaluated for classification accuracy obtained from the various datasets. Numerous statistical parameters such as accuracy, precision, recall, and f-measure are analyzed for the implemented system. The implemented model has achieved higher accuracy 97.8% for multi-class classification.

4.15 Conclusion

In this chapter, the workflow of the deep SQNet model for identifying and diagnosing the coronavirus are discussed. Precisely, the various essential aspects such

Table 4.9: Strength/Weaknesses

Authors	**Model Used**	**Techniques used**	**Strength**	**Weakness**
Verma et al. [139]	VGG-16 with SVM	Segmentation and augmentation	Utilized lung segmentation to focus on classification and acheived 98% accuracy.	A relatively small dataset with 470 COVID-19 images, Overfitting problem due to the limited dataset size.
Khan et al. [140]	VGG-16 and VGG-19	Transfer learning, fine-tuning and has acheived 94.8% accuracy.	Effectiveness of transfer learning for small-sized datasets. This approach achieved optimal classification performance in detecting COVID-19	Not much better performance for multiclass classification.
Samir et al. [141]	ResNet18	Data augmentation, transfer learning	Applied transfer learning and leveraging pretrained models for improved performance has acheived 94.1% accuracy.	The Performance was dependent on the pretrained models.
Kulkarni et al. [142]	CNN and Transfer Learning	Image processing, multiclass classification and has acheived 94.8% accuracy.	Compares multiple models to identify the best performance for classification.	Transfer learning models may not be fully optimized for specific datasets.
Mwendo et al. [143]	Various Deep Transfer Learning Models	Transfer learning techniques	Provides a comprehensive review of deep transfer learning applications in detecting COVID-19, pneumonia, and tuberculosis.	Lacks of specific performance metrics for individual models.
Presented System [133]	SQNet	Fine-tuning, data augmentation	Achieved high accuracy for binary and multiclass classification for COVID-19 when compared to other models.	It may require significant computational resources and further validation to ensure its effectiveness across diverse datasets.

as dataset collection, sources, and classification are also discussed in this chapter. Further, the experimental setup, implementation, prediction model efficiency, and research outcomes have also been described in this chapter.

Chapter 5

Research Methodology for Omicron

5.1 Introduction

This chapter represented the another research work that helps for the identification of other airborne disease Omicron which is variant of coronavirus. For such purpose, an efficient machine learning prediction model has implemented for identifying and diagnosing the omicron patient suffering from such disease.

5.1.1 Background

On November 11, 2021, the omicron (OMCRN) is recognized as a variant of coronavirus (CorV) as the first case was detected in Botswana. The OMCRN cases have also been found in the United Kingdom, Italy, and Belgium. A high alert is announced in Europe as another sequenced case has been found in Hong Kong and South Africa. The OMCRN is a heavily CorV variant also known as B.1.1.529 and On November 26, 2021 the World Health Organization (WHO) listed by as a variation of concern. The WHO announced naming the variant Omicron after Greek Letters to avoid public confusion and stigma. The structure of OMCRN has more mutations on its spike protein. In South Africa, approximately 280 cases

Table 5.1: Symptoms-based Analysis

	Symptoms and their Nominal values	
Severe	**Value**	**Type**
· Shortness of breadth	No, Moderate, Less, High, Very High	Text
· Chest pain	No, Moderate, Less, High, severe	Text
· Loss of speech	Yes, No, Moderate	Text
· Confusion	No, Moderate, Less, High, Very High	Text
Mild		
· Fever	No, Normal, Less, High (Sudden), Severe	Text
· Cough	No, Yes, Less	Text
· Tiredness	No, Yes	Text
· Loss of Taste/ smell	No, Yes	Text
· Irritated Eyes	No, Yes, Less	Text
Less		
· Diarrhea	No, Yes, Less	Text
· Rashes on skin	No, Less, High	Text
· Discoloration of fingers/ Toes	No, Yes	Text
· Sore Throat	No, Yes, Less	Text
· Body pain	No, Moderate, Less, High, Severe	Text
· Head ache	No, Yes, Less, High, Severe	Text

of Corv per day were found before the detection of OMCRN. The United States and Australia have found many confirmed cases of the variation of Corv. The UK Health Security Agency stated that more than 75 new cases had been discovered in England, raising the total number of verified cases. In many nations, the OMCRN is the dominating variant of concern throughout the fourth wave. It is currently the most common Corv variant, causing more than 99% of COVID-19 infections in some states and a massive surge in hospitalizations.The symptoms of OMCRN are divided into 3 categories: Severe, Mild, and Less as shown in Table 5.1.

5.2 Layered Architecture

This section gives a layered architecture of the proposed framework. It comprises different layers namely Data Sensation and Pre-processing Phase (DSAPP), Data Classification Phase (DCP), Data Filtering Phase (DFP), Data Prediction Phase (DPP), and Visualization and Decision-Making Phase (VADMP).

5.2.1 Data Sensation and Pre-processing Layer

This layer is amenable to aggregate data that may be related to health data from the various IoT devices. The proposed system acquires the personal and health related parameters of a person.

Pre-processing - The Data sensation layer collects health parameters such as fever, chest pain, headache, body pain, cough, rashes on the skin, vomiting, nausea, loss of taste, sore throat, irritated eyes, and shortness of breath via smartphone or web-based application. The individual will register themselves via mobile application and give data related to the health attributes in the form of 'Yes' or 'No' and the environmental data are collected from the numerous surroundings environments. A global positioning system is used to determine the exact location of an individual embedded in their smartphone via its navigating tracking system. Table 5.2 represents the category of datasets and their attribute values. Further, the collected data are sent to the fog layer where it will be analyzed for further

Table 5.2: Categorization of Data Sets

Category of Dataset	Attributes	Format
Health Data	Fever, Chest pain, Confusion, and Disorientation of fingers, Appetite, Headache, and Sore throat	Text data
Meteorological Data	Humidity, High(Maximum) and Low (Minimum) temperature	Text data
Location Data	Infectious areas/ Risk suspected regions	Frequency-based non-textual

processing into the possible classification of omicron-infected (IN-OMCR) or uninfected (UIN-OMCR) class. For security concerns, all data can be protected through Secure Socket(SSL) and Transport security Layer(TSL) Protocol. Furthermore, Amazon EC2 can be used for storage of data to increase the efficiency and accuracy. The acquired information is transmitted to the fog layer via fog computing devices for pre-processing and classification purposes. The categorized data are transmitted to the cloud layer for monitoring, in-depth analysis and alerts will send to the concerned stakeholders if any abnormalities found.

5.2.2 Data Classification Phase(DCP)

After the acquisition of health-related parameters from various IoT devices or biosensors, it can pass to fog computing devices for analysis at the fog layer. The DCP layer is responsible for evaluating effective data analysis over such devices. Fog computing devices plays a critical role in real-time data processing by handling computations closer to IoT devices before sending processed data to the cloud.The categorized data are transmitted to the cloud layer for monitoring, for further usage, and in-depth analysis. Moreover, such devices generate on prompt decisions by analyzing the various attributes. Since IoT sensors generate heterogeneous and large amounts of data in real-time applications, Raspberry Pi can be used for data acquisition from such types of devices in the real-time based application scenario. In DCP, fog layer is responsible for data filtering, analyzing, and classification before in-depth processing at cloud. Based on healthcare, the following datasets are formulated by incorporating event-oriented data and data is pre-processed via an efficient parameter extraction mechanism.

1. ***Health-oriented based Data*** - Such data set includes various vital symptoms consisting of Fever, Chest pain, Confusion, Disorientation of fingers,

Headache, and Sore throat. Various human body wearable sensors and biosensors are embedded in the body of an individual to acquire such types of data.

2. ***Meteorological Data*** - Environmental and natural conditions-based data are categorized in this dataset such as high and low temperature and humidity. Numerous sensors like climate and temperature sensors can be placed around surroundings for acquiring data.
3. ***Geographical-based Data*** - It finds the location of risky areas. Global positioning systems (GPS) can be used to collect such types of data.

Moreover, these events-based datasets play a vital role in diagnosing an infected omicron patient.

5.2.3 Data Filtering Phase(DFP)

As stated previously, the primary function of the data classification layer is to classify the datasets based on Weighted Naive bayes(WNB). WNB is a probabilistic classifier that works based on the Bayes theorem for categorization purposes [144]. For further assessment Omi-health Severity Index(OHSI) measure and these datasets are categorized into IN-OMCR(infected) and UIN-OMCR (non-infected) classes. These classes are described as:

1. **IN-OMCR Class** - In such a class, it contains those values which is not lies in between the normal range. In other words, data attributes like high fever, high blood pressure, sore throat, shortness of breadth, and chest pain belong to this class. This category of dataset comprises those attributes that directly or indirectly effects the person health.

2. **UIN-OMCR Class** - In contrast, it contain those data values which lie within the normal limit. These data values are used to determine an individual health condition and non-vulnerable to human health.

5.2.3.1 Mathematically Analyzation

Let D= $\{d_1, d_2, d_3, ..., d_n\}$ be n data points, C = $\{c_1, c_2\}$ be set of different classes. The likelihood function P(y/c_j) and y $\in$ D for given the classes. o The probability of observing a set of characteristics can be calculated as[145]:

$$P(c_j/y) = \arg\max_{c_j \in C} \left[\frac{P(c_j) * \Pi_{j=1}^{2} P(c_j/y_i)}{\Pi_{i=1}^{n} P(y_i)}\right] \qquad (1)$$

In all the denominator is constant so it can be removed from eq(1). Then performance of weighted NB by assigning different weighting to each attribute can be written as:

$$P(c_j/y) = \arg\max_{c_j \in C} \left[P\left(c_j\right) \times \prod_{\substack{i=1 \\ W_i \in (0,1)}}^{n} P\left(c_j/y_i\right)^{W_i}\right] \qquad (2)$$

A positive number which is given to all attributes known as weight w_i. WNB is used for determining the probabilistic value of each parameter, based on the vulnerability scale known as Omi-health Severity Index.

Definition 1: Omi-health Severity Index (OHSI)

The probabilistic value is used to measure the OMCRN-infected health severity of a person at the time δt. If the probabilistic value of a person lies in between the the threshold value given by an expert, then it is safe, otherwise, an alert message is to be generated to the stakeholders.

5.2.3.2 Time-series Based Data Extraction

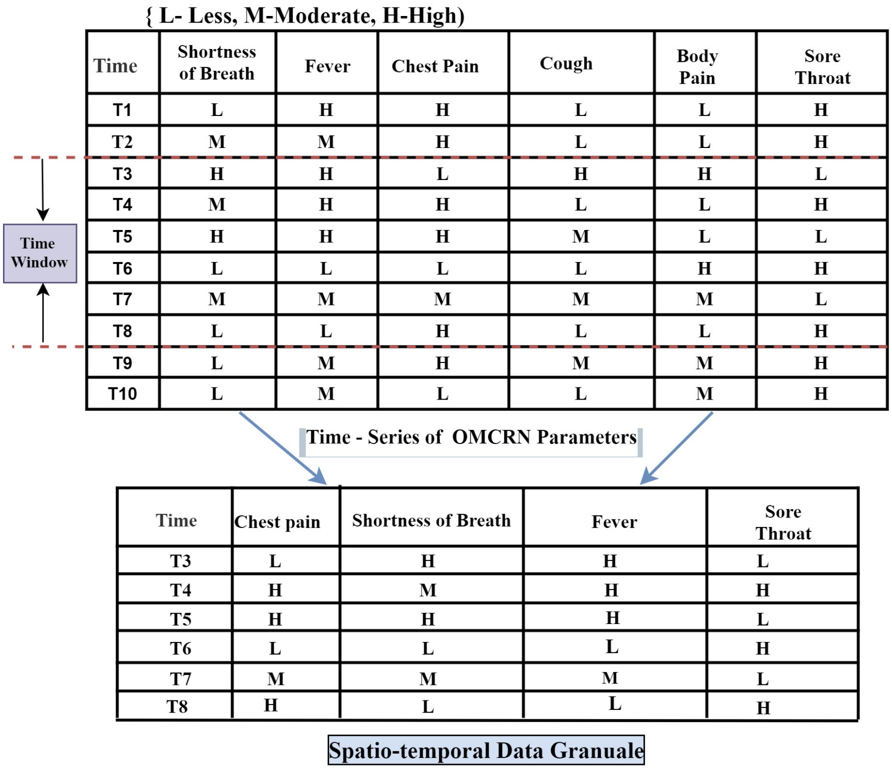
After filtering the attributes, the spatio-temporal (SPT) based data mining technique is used to extract the OMCRN health data at time window δt [46]. Both geographical and temporal data can help to improve the performance accuracy and predict the probability of an infected person on time. The approach which is used for SPT-based feature extraction called as SPT granularity.

Definition 2: **Spatio - temporal Granularity** :
The temporal-granularity G_T and spatial(S) elaboration over the temporal domain of $S_{temporal}^{G_T}$ = $<G_{d_i}$,$ES_j >$ ∀ i=12,3....,m , j=1,2,...,k. Each data value represented by d_i of kth event at time window δt where $t_j \in$ [t,t+δt][146].
Figure 5.1 represents the SPT Granularity at time window δt. The information that is retrieved based on SPT must be stored as instances in the data sets[147].

For n attributes the omicron vulnerability score will be evaluated. Algorithm 1 (Table 5.3) defines the step-by step procedure to evaluate the OMCRN scale of vulnerability (Osov). If the valve of O_{sov} value lies with in limit it means person belong to safe UIN-OMCR class otherwise a person is not safe and belong to IN-OMCR class. So, a warning alert message will be generated(as explain in algorithm 2).

5.2.4 Data Prediction Layer(DPP)

The main goal of DPP is to determine the prediction analysis effectively. The prediction layer is utilized for the identification of a person suffering from OMCRN and it used temporal data for predictive analysis. For such purpose, an enhanced temporal data-based recurrent neural network (ERNN) in collaboration LSTM is utilized. ERNN is mainly made of input components, hidden units with neurons, and one output unit. It consists further comprises Monitoring, Training, and Prediction phases which are ahead:



Time	Shortness of Breath	Fever	Chest Pain	Cough	Body Pain	Sore Throat
T1	L	H	H	L	L	H
T2	M	M	H	L	L	H
T3	H	H	L	H	H	L
T4	M	H	H	L	L	H
T5	H	H	H	M	L	L
T6	L	L	L	L	H	H
T7	M	M	M	M	M	L
T8	L	L	H	L	L	H
T9	L	M	H	M	M	H
T10	L	M	L	L	M	H

Time	Chest pain	Shortness of Breath	Fever	Sore Throat
T3	L	H	H	L
T4	H	M	H	H
T5	H	H	H	L
T6	L	L	L	H
T7	M	M	M	L
T8	H	L	L	H

FIGURE 5.1: Spatio-temporal Data Granularity

5.2.4.1 Monitoring Phase

In this phase, ERNN will assign different values to each attribute and then allocates weight on the importance of each attribute. On the basis of training, such phase stores its significant parametric value that can be used for prediction at the fog layer.

Let R_t=$\{r_1, r_2, r_3, .., r_n\}$ be the n input data, where $r_t \in \mathrm{R}^k$ of the weighted matrix W_{ht} at time t. The hidden layer HD_t represented the system state memory at time t as:

$HD_t = \psi_{HD}(\sigma_t)$ where

σ_t=$\mathrm{W}_h R_t + W_{hh}\ hh_{t-1} + bs_0$ (3)

Here $\mathrm{W}_h \in \mathrm{R}^{dxd}$, $W_{hh} \in \mathrm{R}^{dx_k}$, $bs_0 \in \mathrm{R}^d$ updated during training phase. σ and $\psi(.)$ are nonlinear activation functions.

Table 5.3: OMCRN Severity of Vulnerability determination

Algorithm 1 Vulnerability Scale value determination procedure

```
Input: consider $D_i = \{d_1, d_2, d_3..... d_n\}$ be the data points for associated OHSI values at various parameters (prior probability of OMCRN-related attributes) and $O_{w1}$, $O_{w2}$, $O_{w3}$,...,$O_{w_m}$ are associated weights.
Step 1. Initialize $O_{sov}$ to NULL
Step 2. Compare each data value with the predefined threshold value.
Step 3. if $OHSI_{d1}$ > threshold value Then add $O_{w1} * OHSI_{d1}$ to Osov.
Step 4. if $OHSI_{d2}$ > threshold value Then add $O_{w2} * OHSI_{d2}$ to Osov.
Step 5. if $OHSI_{d3}$ > threshold value Then add $O_{w3} * OHSI_{d3}$ to Osov
Repeat for n data values
Step 6. If $OHSI_{dn}$ > threshold value Then add $O_{wm} * OHSI_{dn}$ to Osov.
Step 7. $O_{sov} = O_{w1}*OHSI_{d_1}+O_{w2}*OHSI_{d_2}$ +.......$+O_{wm}*OHSI_{dn}$
```

5.2.4.2 Training Phase

The training phase helps to diminish the error till the desired tolerance which was produced during monitoring phase. To achieve such goal, the ERNN consists of two processes feed-forward and to update the weight of each parameter, the back-propagation process is used.

LSTM-ERNN Model

In ERNN, when data is transmitted backwards, the problem of vanshing gradient appered. To solved the gradient problem, the LSTM is used which has the ability of faster convergence and diminish the problem of exploding gradients.

The proposed LSTM-ERNN model consists of hidden layers as a memory cell for storing information for a long period of time and the rest of the architecture is similar to existing RNN. Figure 5.2 depicts the individual neuron structure of the LSTM-ERNN with memory cell(c_t) at time t and hh_t represents the hidden state. LSTM cell c_t include various gates namely Input (I_G), Forget (F_G), and Output (R_G) gate.

To enhance the prediction accuracy results, the proposed model used the LSTM. At each time step t, inputs (r_t , hh_{t-1}, c_{t-1}) are sent to three gates which generates

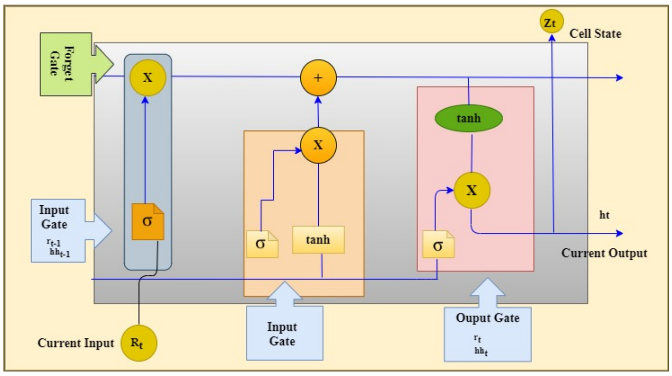


FIGURE 5.2: Schematic view of LSTM-ERNN

signals that can be evaluated using following equations:

$F_G^t = \sigma(W_f \times [hh_{t-1}, r_t] + b_f)$ (4)

Input Gate (I_G)

I_G is responsible to determine and add new information through activation functions. It is used to generate parameters and to update information by using the following equations[145].

$S_I^c = tanh(W_i \times [hh_{t-1}, r_t] + b_c)$

$S_R^c = f_t * S_R^{c_{t-1}} + I_G^R * S_I^c$ $I_G^R = \sigma(W_i * [hh_{t-1}, r_t] + b_i)$ (5)

where,

S_I^c is current input. I_G^R denotes the result of I_G W_i and b_i denotes the independent weight and bias of I_G respectively.

Output Gate (R_G)

It calculates the final output of the current state to complete the forward propagation calculation of the network by using the formula:

$R_G = \sigma(W_f \times [hh_{t-1}, r_t] + b_f) * tanh(S_R^c)$ (6)

where $tanh(r) = \frac{1-e^{-2r}}{1+e^{-2r}}$ (7)

5.2.4.3 Prediction Phase(PP)

This layer is utilized for the predictability of O_{sov} in an associative time window is calculated. Mathematically, it is expressed as:

$$P(O_{sov}) = \sum_{i=1}^{t} \alpha_t hh_t \qquad (9)$$

where, $\alpha_t = \frac{exp(S(hh_t,\gamma))}{\sum_{t=1}^{T} exp(s(hh_t,\gamma))}$ such that $S(hh_t, \gamma)$ is a mapping function, α_t is a normalized value that refers to the whole network at time t.

5.2.5 Visualization and Decision Making Layer(VADMP)

This layer is utilized as a presentation layer for a person in order to facilitate the provision of healthcare services effectively. The primary function of this layer is to provide the user to depict the O_{sov} outcome via a hand-held device or a Liquid crystal panel or a smartphone for efficacy. In addition, to display the parameters numerically, visualization is incorporated with the SPELLMap technique. It integrates the spatial and temporal data to perform analysis. In order to enhance the visualization, the different colour coding schemes is used for the representation the effect of omicron. In the presented study, the U-matrix depicted method is used for prediction indication of the proposed model. It is depicted as follows;

O_{sov} (**Low**) – > **Green Colour**
O_{sov} (**Medium**) – > **Yellow Colour**
O_{sov} (**High**) – > **Red Colour**

The process of the colour coding with U-matrix mapping is shown in Figure 5.3. The red color indicate a high O_{sov} value whereas green color indicates a low O_{sov} value. Yellow colour indicate the medium O_{sov} value.

It's important to note that data can be transmitted between the fog computing device and the display screen via a variety of wireless communication protocols such as Bluetooth, Wi-Fi, or ZigBee. In addition, along with the vulnerability prediction technique, the alert message procedure has included to increase the facility

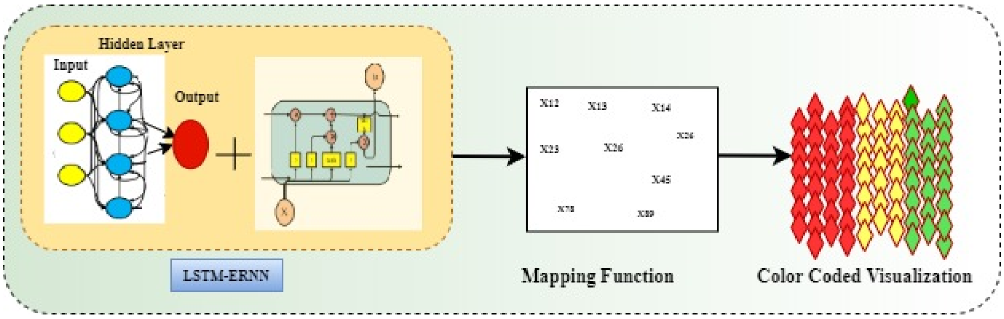


FIGURE 5.3: LSTM-ERNN with U-matrix

TABLE 5.4: Steps for generation of alert signal

Algorithm 2. Warning Signal Generation Procedure

```
Input: Current category(catg) of a person, parameters, Probability value of numerous attributes, and Pre-defined threshold values.
Step 1. Take present category(catg),various OMCRN health parameters at the present time window of a patient.
Do
For j = 1 to n parameters
Step 2. If (Attribute (j) > threshold-value)
2.1 If (catg == IN-OMCR ) Then
2.2 Person is unsafe and go to step 4.
Else
Step 3. If (catg == UIN-OMCR)
Step 4 Calculate Osov.
4.1 If P(Osov) > Pre-defined-threshold-value Then
4.2 Send warning message to a patient or caretakers or hospitals.
End Do
Step 5. Goto Step 2.
Step 6. Present output to SPELLMap (for display).
```

for an individual suffering from omicron so that action can be taken promptly. Algorithm 2 (Table 5.4) shows the visualization procedure including to send an alert generation.

5.3 Predictive Mining

Predictive mining has been able to explore the knowledge available in the healthcare domain to improve decision-making and anticipate future trends or events more accurately [148]. It helps in designing of a prediction model to support decisions. The main task of this approach in the classification involves assigning a category to a new observation or instance after the model has learn from the training set. It is also known as supervised learning. After training, the classification model performs testing on the data. The predicted values are compared for knowing values i.e., the accuracy of classification is measured. Accuracy is the evaluation criteria of a classification model wherein the percentage of observed classification values are compared with the actual classification.

5.4 Experimental Setup

This section describes the experimental setup for the proposed system. For enhancing the another research work, the suggested prediction model for Omicron was implemented. For this purpose, the experiment was performed in MATLAB on Intel Core i7 Quad-core CPU with 3.2@GHz processor and RAM of 4 GB for training and testing. Further, the numerous evaluation metrics are compared with state-of-the-art models of the proposed system.

5.4.1 Dataset Source

The data is to be collected from the open-source repository GitHub[1]. It consists of symptom-based data for the two possible classes (i) UIN-OMCR(non-infected) and IN-OMCR (Infected) class. In the given dataset, various parameters such as fever, sore throat, chest pain, sex, cough, headache, tiredness, and shortness of breath are considered. The collected dataset consist of 9525 disease-infected cases, 2851 healthy, and 2500 non-confirmed cases.

---

[1] source:https://github.com/nshomron/covidpred

5.5 Results and Performance Analysis

5.5.1 Training and Testing Implementation

This subsection how a dataset is split into training and test dataset for building prediction model for Airborne disease. The investigator has split dataset into two parts in the ratio of 80:20. The first part is a training set while the second part is of test dataset which has shown in Figure 5.4.

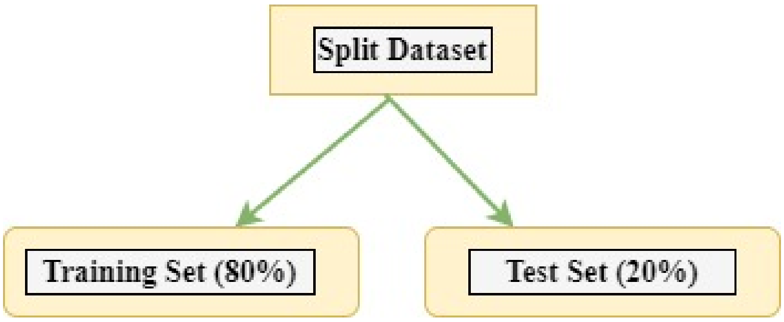


FIGURE 5.4: Dataset distribution

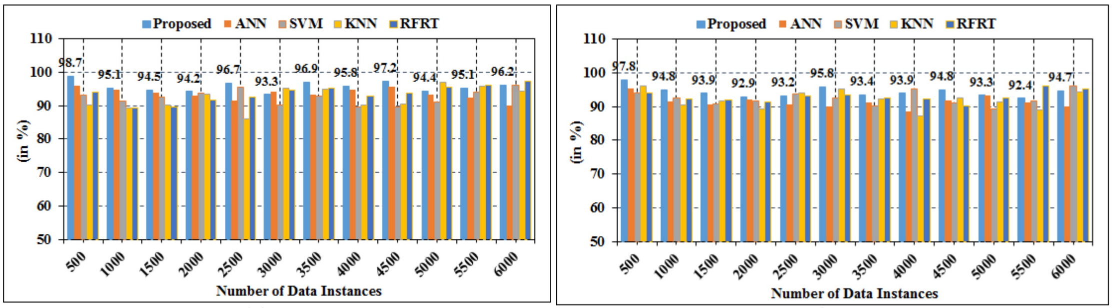
5.5.1.1 Performance analysis of Classification

The presented model is evaluated using 4 quantitative statistical metrics in terms of Accuracy(ACR), Precision(PRC), Recall(REC), and F-measure(FM) to evaluate the classification performance of the proposed system. The MatLab environment is used for experimental purposes. Table 5.5 depicts the classification analysis performance of the proposed model. The experimental result exhibits that the suggested model has achieved higher accuracy when it is compared to other classifiers models such as Support vector machine (SVM), ANN, K-Nearest Neighbor (KNN), and Random Forest Tree (RFRT) [149][150].

.

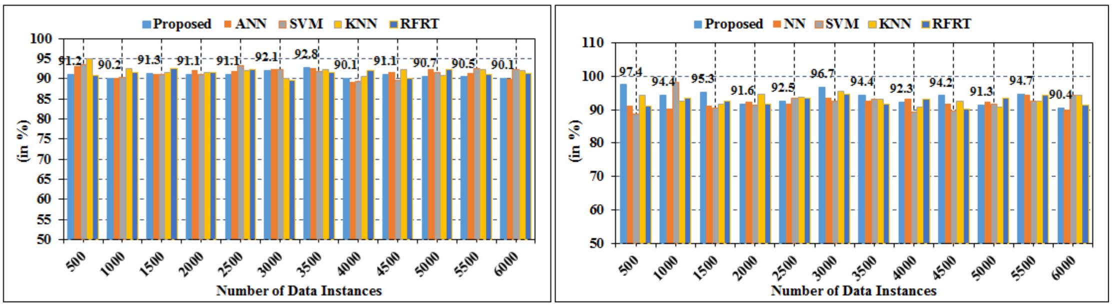
Table 5.5: Classification Performance

Performance Parameters	ACR (in %)	REC (in %)	PRC (in %)	FM (in %)	MCRC
ANN	95.8	95.2	92.5	93.5	0.95
SVM	96.1	96.1	92.3	94.3	0.95
KNN	96.9	95.4	92.1	95.6	0.96
RFRT	97.3	96.2	92.2	94.7	0.97
Proposed	**98.7**	**97.8**	**92.8**	**97.4**	**0.99**



a) Accuracy

b) Recall



c) Precision

d) F-measure

Figure 5.5: Efficacy of Proposed Model (a) Accuracy (b) Recall (c) Precision (d)F-measure

1. Figure 5.5(a) depicts that the proposed model has achieved higher accuracy of 98.7% as compared to ANN (95.8%), SVM (96.1%), KNN (96.9%), and RFRT 97.3% respectively.

2. Figure 5.5(b) depicts the presented model has obtained a recall value of 97.8% which is much better than other models such as ANN consists of 95.2%, SVM consists of 96.1%, KNN consist of 95.4%, and RFRT 96.2% respectively.

3. Moreover, Figure 5.5(c) depicts that the proposed model has obtained a precision value of 92.8%, that is better than ANN, RFRT, and KNN.

4. Furthermore, Figure 5.5(d) has shown value of FM 97.4% of the proposed model which is higher as compared to SVM (94.3%), ANN consists of 93.5%, and RFRT consist of 94.7%.

As a result, it can be concluded that the classification performance of proposed model has achieved more efficiency than others.

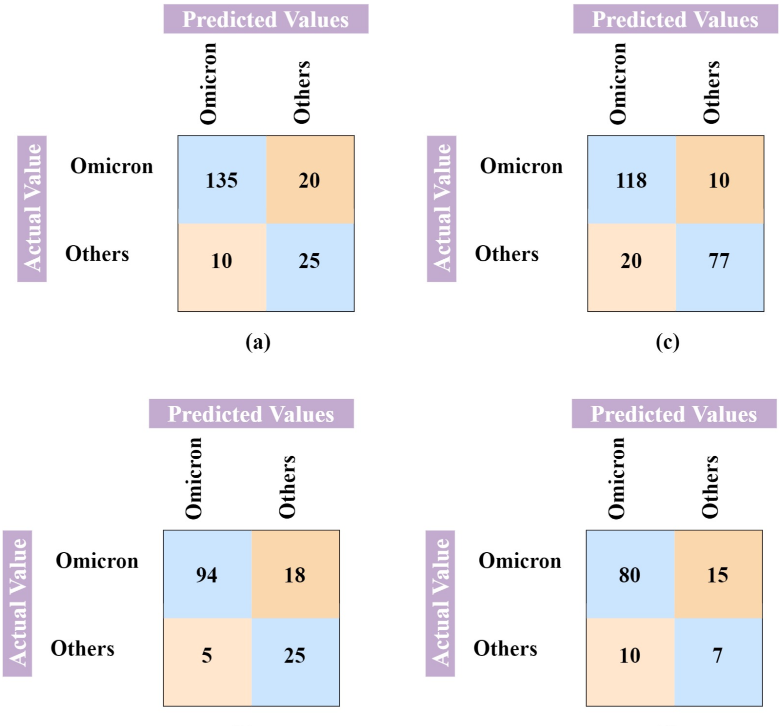
Confusion Matrix

The confusion matrix shows the relationship between true positive and true negative. It provides the ability for understanding the flawed of proposed model in addition to the TP and FP. Figure 5.6 a,b,c and d depicts the confusion matrix of SVM, ANN, and the proposed model respectively.

Training Accuracy and Loss Curve

In LSTM-ERNN the proposed modal consist of 12 neurons at the input layer and the 1 neuron at output layer during training and testing. The Fully Connected Layers of the ERNN model contain 15 nodes in the 1st hidden layer, 10 nodes in the 2nd hidden layer. The ADAM optimizer was used for the ERNN model along with the activation functions. Figure 5.7 represented the validation loss and accuracy of every epoch during training and testing of the proposed model. The learning rate of the implemented model was 0.001 and it achieved 0.1 dropping out rate for regularization of the processed information.

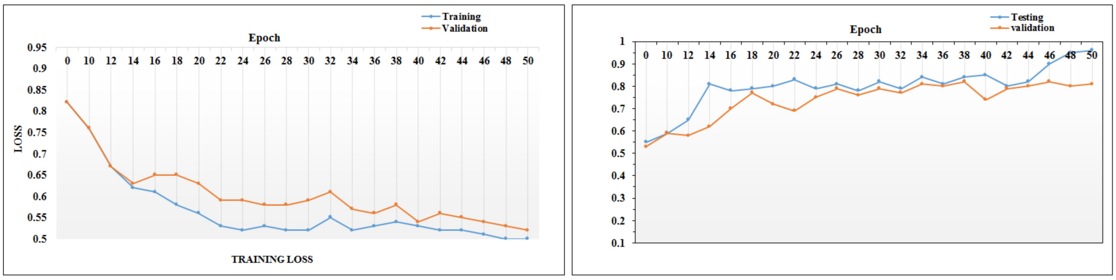
To explore the performance analysis, the presented model is compared with another classification models at different instances. Table 5.6 depicts the performance of presented model and also shown the comparison with other models in terms of



(a) (c)

(b) (d)

FIGURE 5.6: Confusion matrix of a) SVM b) ANN c) RFRT d)Proposed



a) Model Loss Curve

b) Model Accuracy Curve

FIGURE 5.7: Proposed model Loss(Left) and Accuracy(Right) Curve

various parameters at different instances.

1) As it can be seen in Figure 5.8(A, B,C, and D), the reported value of recall is increasing as the data instances are increasing. Figure 5.8 (A) represented that at 1500 data instances highest recall value was 95.7%, at 4500 instances recall value increased to 96.8% of the proposed classifier.

2) Moreover, it can be seen for other metrics such as precision and f-measure. Figure 5.8(E,F, G, and H) have shown the different values the precision values at 1500, 2500, 5500, and 6000 are 90.7%, 91.1%,92.4% and 92.8% respectively. It can

be noted that the precision of the proposed classifier has achieved a higher value of 92.8% as well as instances increases.

3) Similarly, Figure 5.8 (I,J,K, and L) have shown the obtained value for f-measure of the proposed classifier that has achieved a higher value of 97.4% as well as instances increases.

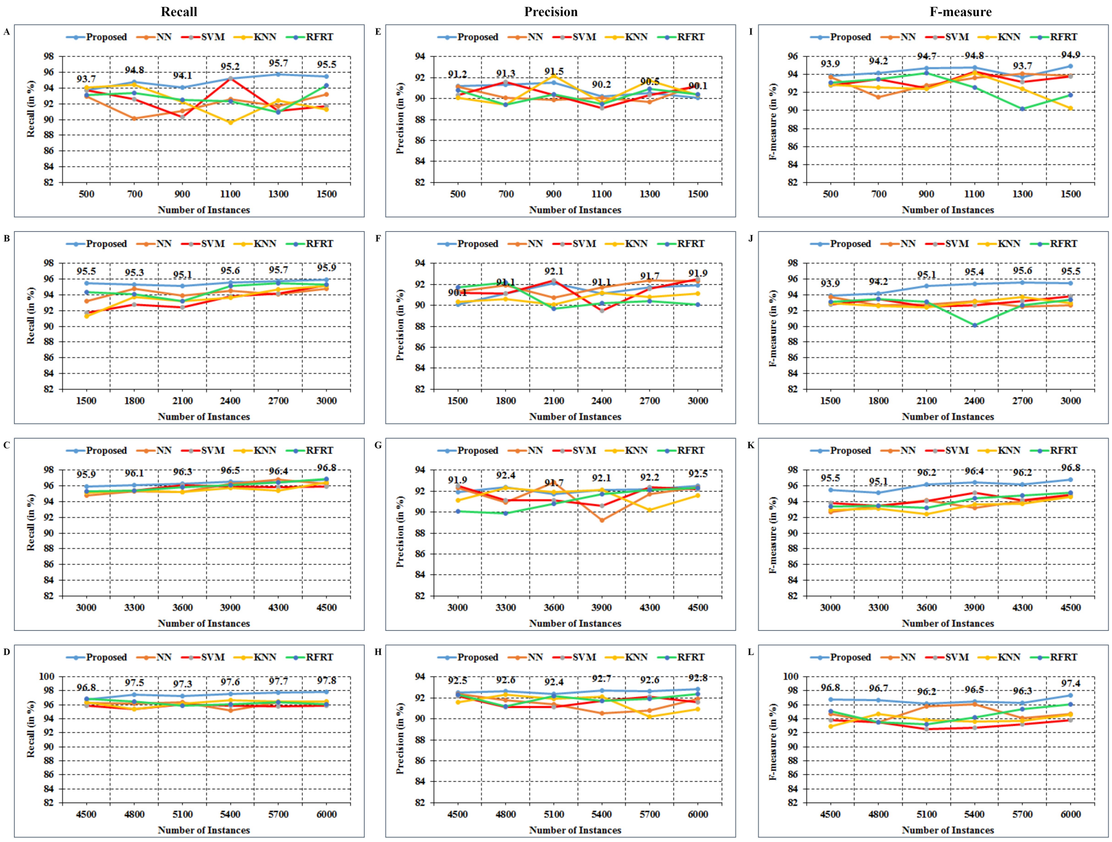


FIGURE 5.8: Performance efficiency at different instances

5.5.1.2 Prediction Efficiency

The performance of the proposed model is evaluated using various statistical parameters such as precision etc. In addition, to enhance the performance of proposed model, Root Mean Square error (RSE) and Squared Pearson correlation coefficient (SPCRC) are computed for error rates in prediction. As demonstrated

Table 5.6: Classifications Efficiency

Model	**Proposed(in %)**			**SVM(in %)**			**RFRT(in %)**		
No. of Instances	**REC**	**PRC**	**FM**	**REC**	**PRC**	**FM**	**REC**	**PRC**	**FM**
500	92.4	90.2	91.3	90.3	90.4	89.1	90.2	91.2	92.4
1500	93.9	90.7	92.5	90.8	91.6	90.2	91.8	91.7	90.2
2500	93.2	91.1	91.6	92.4	90.8	92.5	93.4	90.2	91.7
3500	93.4	91.7	94.4	93.7	90.4	93.2	92.5	91.3	92.6
4500	94.6	92.1	95.3	93.2	91.1	93.4	94.1	92.5	91.8
5500	95.7	92.4	96.3	93.5	90.5	93.9	93.7	91.1	93.3
6000	97.8	92.8	97.4	94.1	91.7	94.3	94.2	91.7	94.2

in Table **??** the proposed framework achieves least error in terms of RSE amd SPCRC. The suggested model is contrasted with three models, namely Artificial Neural Network (ANN), Support Vector Regression (SVR), and Random Forest Regression (RFR).

1) As it can be noticed that in Table 5.6 the reported value of presented model consists of precision 92.8%,recall value consist of 97.8% respectively at different instances. Thus, it can be seen that the classification efficiency of the proposed classifier has a higher as compared to SVM and RFRT.

Table 5.7: Analysis using Statistical Errors

Models	**SPCRC**	**RSE**
ANN	0.81	1.8
SVR	0.91	1.5
RFR	0.73	1.7
Proposed	0.51	1.1

2) In addition, it can be seen in Figure 5.9, the proposed model has obtained less error in terms of RSE(1.1±0.20%) and SPCRC (0.51±0.42%) as compared to ANN, SVR, and RFR. As a result, the implemented model has enhanced prediction results with a tolerable error rate.

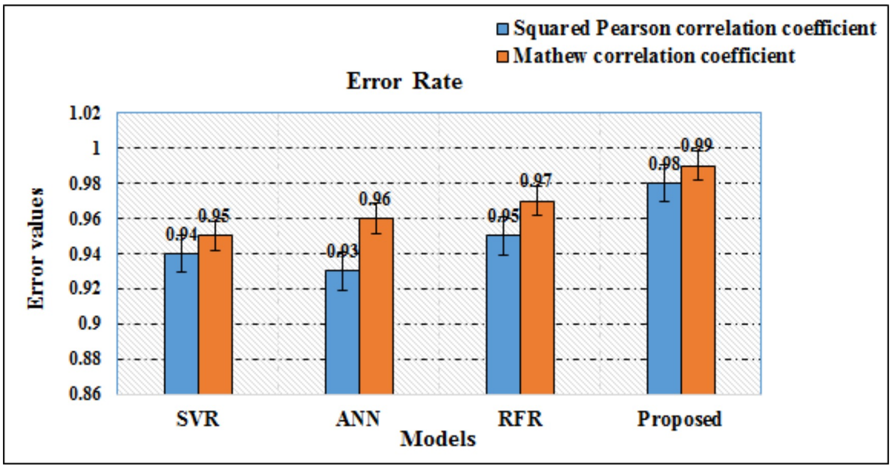


FIGURE 5.9: Error Rate of Presented Model

5.5.1.3 Prediction Time Efficiency

In this subsection, prediction time efficiency is measured incorporate with the latency time to explore the efficiency of the proposed model for alert generation. It is the time difference between performance analysis(PR_{anly}) and generation of respective alert(GEN_{alrt}) to display result.

It can be calculated by using following formula:

$L_{Time} = PR_{anly}$ - GEN_{alrt}

The Figure 5.10 represented the prediction time efficiency, which has been calculated by the utilization of latency time. Moreover, it can be observed that the evaluated time of the proposed prediction model is less as compared to other models like SVR, ANN, and RFR.

In Figure 5.11, it can be seen that the prediction performance time of the proposed model is 0.51s which is less as compared to other ANN, SVR, and RFR. Figure 5.11 represents that the evaluation time for creation of warning alert message. It can be concluded from the analysis results that the presented approach is cost-effective and helps for combating the AB disease.

At the end of this chapter, it can be observed that the presented prediction model has been performed efficient in terms of statistical evaluation metrics when compared with numerous SOA of other models.

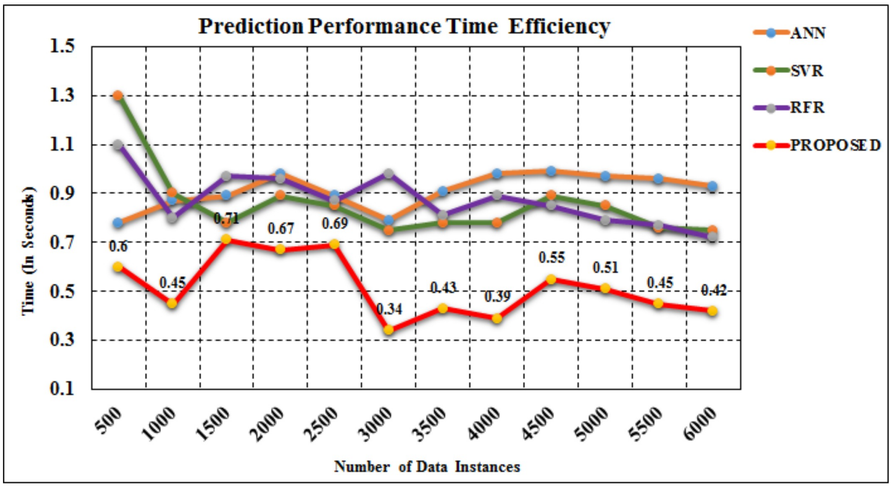


FIGURE 5.10: Prediction time Efficiency

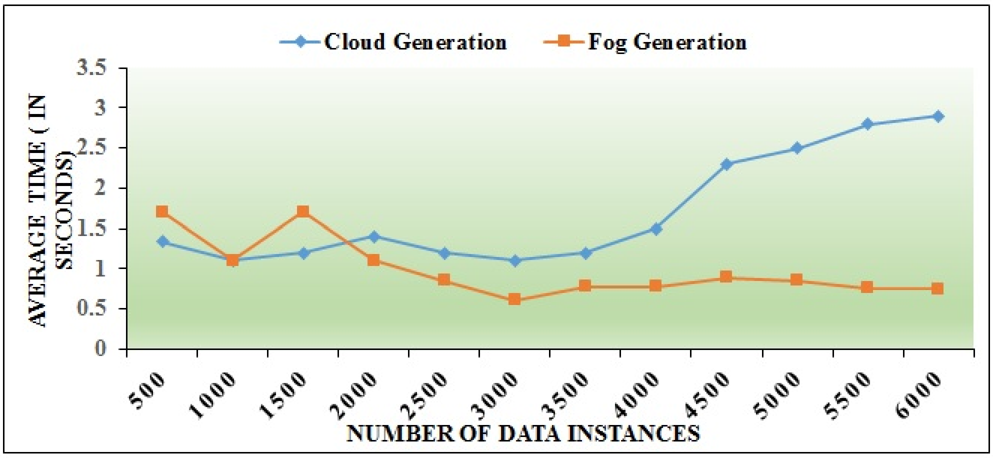


FIGURE 5.11: Alert Generation time

5.5.2 Comparison with Other Models

The outcomes received from LSTM-ERNN is compared with other models such as CNN, Stacked-LSTM, and Autoregressive integrated moving average(ARIMA) which is a time series model as shown in Figure 5.12. It can be observed that ARIMA performs better in predicting the OMCRN as compared to stacked-LSTM. However, the implemented model perform better than the other models as depicted in Table 5.8.

TABLE 5.8: Comparative Prediction performance analysis with other models

Models	**MCRC**	**NPV**	**PRC**	**ACR**
Stacked-LSTM	0.92	0.95	0.94	0.95
CNN	0.91	0.96	0.93	0.94
ARIMA	0.92	0.94	0.92	0.96
Proposed	0.93	0.97	0.96	0.94

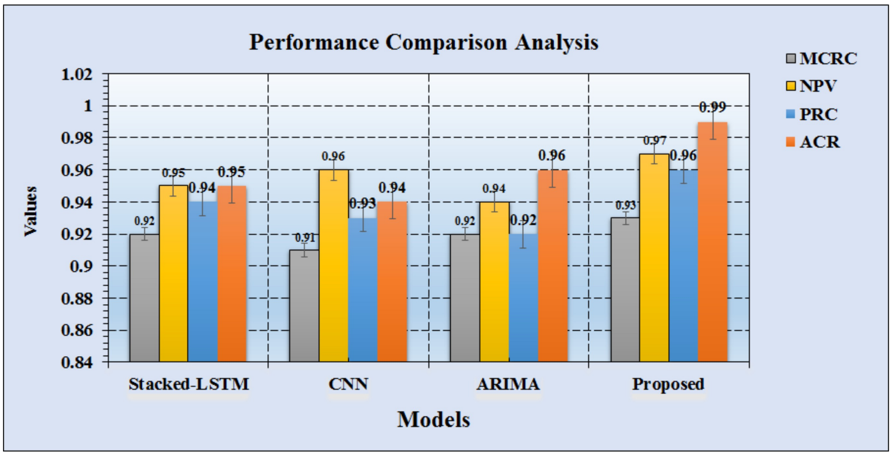


FIGURE 5.12: Comparison with Other models

5.5.3 Comparative Study With Others Recent Work

Using machine learning algorithms has gained progress in the early diagnosis of such types of illnesses. DL techniques such as Recurrent neural networks and convolutional neural network techniques outperform for diagnosing ABDs in terms of statistical parameters. Li et al. [135] present a deep-based architecture for identifying coronavirus patients. The experiment results have shown the presented model has obtained 89.1% accuracy via transfer learning. Jaiswal et al. [151] presented a DenseNet201 deep learning model for the classification of the Coronavirus variant using transfer learning. The experimental results have shown that the accuracy obtained by the presented model was 97%.

In the current thesis, the classification algorithm and LSTM-ERNN prediction Model has presented as a fast, reliable, and efficient technique to diagnose the OMCRN. Furthermore, the current work has shown the presented classification model has achieved higher accuracy (98.7%) when compared to other models such as ANN, SVR, and Random forest tree. Table 5.9 depicts the comparison of the presented work with other works.

By comparative studies, it can be concluded that the presented model achieves 98.7% accuracy, outperforming existing studies such as DenseNet201 (97%) and CNN-based models (96.8%). The use of LSTM-ERNN enhances real-time prediction performance, leading to better results and can provide a faster response time in healthcare decision-making.

5.6 Discussion

In this work, a a predictive model has implemented for effectively and accurately diagnosing, monitoring and predict the omicron airborne disease. by using obtained dataset. A prediction model LSTM-RNN has been implemented to diagnose, monitor and predict the omicron (coronavirus variant) cases. By experimental results, it can be observed that the proposed model has obtained the higher accuracy 98.7% as compared to other prediction models. Furthermore, the performance evaluation results have shown that the predictive model has achieved better results in terms of precision, f-measure, and recall when it compared to other state-of-the-art models.

5.7 Conclusion

This chapter discussed the distinct layers of the model. Each layer has performed their task to achieve objectives. Precisely the various significant aspects such as data collection, data classification, spatio-temporal based feature extraction, and prediction are discussed in this chapter.

Table 5.9: Comparison with Recent works

Authors and References	**Model/Technique**	**Temporal Analysis**	**Key Features**	**Accuracy**
Song et al. [98]	Deep Learning(CNN model)	No	COVID-19 identification via transfer learning	89.1%
Panwar et al.[119]	CNN based deep learning model	No	X-ray-based COVID-19 fast detection	97.1%
Jaiswal et al. [151]	Transfer Learning, DenseNet201	No	Classification of COVID-19-infected patients	97.1%
Manocha et al. [152]	Machine learning	No	Prediction framework for COVID-19 detection	92.8%
Hertel and Benlamri [153]	Deep learning based CON-SNET model	No	Omicron variant detection, real-time classification, decision-making phase	96.8%
Saputra and Bernarte [154]	Machine Learning	No	Naive Bayes Spread prediction of the Omicron variant in Indonesia	84.4
Presented System [155]	ML and Deep learning Algorithms	Yes	Omicron variant detection, decision-making phase	98.7%

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

With the exponential development and intensification of the process of global integration, public health is gaining attention nowadays. The airborne diseases have increased the mortality rate day-by-day which is primary challenged problem all over the world. It is a big task for the healthcare industry to diagnose the disease timely. Nowadays, The prevalence of airborne diseases has become a global health concern. Last few years, airborne diseases posed a great threat across the world. In 2020, the WHO declared the pandemic COVID-19 as an airborne disease that can be spread by tiny pathogens in the air via directly or indirectly. The pandemic COVID-19 badly influenced the nation socially and economically. Hence, such type of diseases have been badly affecting the individual health, development of the country, and also increasing the mortality rate at alarming rate. Most of the researchers have worked on such diseases but still there is working going for the detection of such diseases at an early stage. There is a huge potential in the area of early detection of airborne disease are required. The investigator has found that there is no platform available for diagnosing and identifying the coronavirus at an early stage. Therefore, the current study motivated to develop of a prediction model that enables to predict the coronavirus ad its variant at an early stage as well as accurately classifying the same so that a decision-making action or treatment can be taken on time. Finally, this chapter provides an overview of the work done, summarizing the work carried out, making overall conclusions, and identifying future directions for research and development. This chapter provides the overview of the presented work which can be concluded as;

Chapter 1 described the outline of presented study. The aim of study, problem statement, motivation, objectives, research gaps, and workflow of proposed methodology are discussed in this chapter. **Chapter 2** described the various existing techniques which are reviewed for diagnosing the airborne diseases. The comparative study of the existing techniques are analyzed and research gaps are presented in this chapter to achieve the first objective. **Chapter 3** provides the problem formulation and deep learning / machine learning methods to achieve the objectives. **Chapter 4** provides detailed information about first research work i.e related to identification of coronavirus. Precisely the various significant aspects such as data collection, data classification, data mining, feature extraction methodology, and prediction are discussed in this chapter. In this chapter, the dataset collection and classification, dataset sources are discussed. In this chapter, outcomes concerning classification efficiency, prediction efficiency, predictive effectiveness, prediction performance analysis are discussed in detail. **Chapter 5** provides detailed information of second research work i.e related to Omicron. Precisely the various significant aspects such as data collection, data classification, data mining, feature extraction methodology, and prediction are discussed in this chapter. It also explains the sources from where data has been collected. Result outcomes, classification efficiency, prediction efficiency, predictive effectiveness, prediction performance analysis are discussed in this chapter. Chapter 4 and 5 have achieved the objective two and three of the current research work.

The proposed system consists of a layered approach to monitor and prevent the spread of the airborne diseases coronavirus and omicron. In the presented work, a CNN-based SQNet prediction model has been presented for the prediction of coronavirus.

In addition, to classification and predictivity the omicron cases, a LSTM-ERNN prediction model has been evaluated. To validate the model, a large number of experiments have been conducted and the results have compared to other state-of-the-art prediction models. Additionally, to evaluate the performance analysis of the proposed model several statistical measures such as accuracy, precision, recall, and f-measure are evaluated.

The main key aspects of the implemented system for coronavirus is as follows:

1. The average predictive accuracy achieved by SQNet model has (97.8%) for four-class and 98.4% for three-class classifications for diagnosing the coronavirus respectively which has higher as compared to other models.
2. The transfer learning approach is utilized to solve the problem of imbalance data and overfitting.
3. The experiment was performed on 5-fold cross-validation for the obtained datasets with an acceptable dropout rate.
4. The suggested model also includes the analysis of the f-measure that has achieved a higher value of 95.12% comparative to other SOA approaches.
5. In addition, the specificity analysis of the predictive model has consists higher value of 93.65%.

Moreover, to enhance the research work for other ABD, the LSTM-ERNN prediction model has also been conducted for monitoring and diagnosing the omicron. The main key aspects of the implemented system for omicron is as follows:

1. The proposed prediction model has achieved 98.7% accuracy for identifying and predict the omicron over spatio-temporal data mining.

2. Moreover, the LSTM-ERNN prediction model has also been conducted for monitoring and diagnosing the omicron. The presented model has evaluated by various statistical metrics such as precision, recall, and f-measure. The experimental results have shown the recall value consists of 97.8%, the value of precision 92.8%, and f-measure consists of the value 97.4% of the presented model.
3. Moreover, the prediction efficiency of the predictive model has evaluated for Root Mean Square error and Squared Pearson correlation coefficient. The outcomes have shown the proposed model has obtained less error.
4. Additionally, an alert message is to be generated to the stockholders if any abnormality found.

Therefore, it is concluded that the proposed system has much better for diagnosing the ABDs.

6.2 Future Directions

Research is an evolutionary process, and the completion of one research problem is the beginning of many upcoming works. The outcomes of the thesis pave the way for future research and provide a researchers and society to make well-informed suggestions in the field of the healthcare system. The following section provide the various future direction in the domain of healthcare system;

1. To successfully address real-world challenges in healthcare and diagnostics, an improved model and also conducted for other airborne disease in the future.
2. The proposed model could be implemented for large dataset to enhance the accuracy and for making the model more robust.

3. The primary aim of the current research is to create an effective model for combating such diseases. In order to demonstrate a real-world based application, Such approaches will be investigated in more detail for other airborne diseases also.

4. As the security concern, data security and ethical data is another important domain for exploration.

5. The outcomes of this study might help researchers for monitoring and diagnosing the airborne diseases on timely.

6. This study also helpful for the society as it provides machine learning based model to prevent the spread of airborne diseases and identify such types of disease timely which could help to provide a healthy life to an individual.

REFERENCES

[1] R. P. Clark and M. L. de Calcina-Goff, “Some aspects of the airborne transmission of infection,” *Journal of the Royal Society Interface*, vol. 6, no. suppl_6, pp. S767–S782, 2009.

[2] S. Tiwari, R. K. Singh, R. Tiwari, and T. N. Dhole, “Japanese encephalitis: a review of the indian perspective,” *The Brazilian Journal of Infectious Diseases*, vol. 16, no. 6, pp. 564–573, 2012.

[3] A. M. C. Thar, K. T. Wai, A. D. Harries, K. L. Show, L. L. Mon, and H. H. Lin, “Reported measles cases, measles-related deaths and measles vaccination coverage in myanmar from 2014 to 2018,” *Tropical medicine and health*, vol. 48, pp. 1–11, 2020.

[4] S. K. Bhardwaj, N. Bhardwaj, V. Kumar, D. Bhatt, A. Azzouz, J. Bhaumik, K.-H. Kim, and A. Deep, “Recent progress in nanomaterial-based sensing of airborne viral and bacterial pathogens,” *Environment international*, vol. 146, p. 106183, 2021.

[5] K. E. Jones, N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak, “Global trends in emerging infectious diseases,” *Nature*, vol. 451, no. 7181, pp. 990–993, 2008.

[6] N. Wilson, A. Norton, F. Young, and D. Collins, “Airborne transmission of severe acute respiratory syndrome coronavirus-2 to healthcare workers: a narrative review,” *Anaesthesia*, vol. 75, no. 8, pp. 1086–1095, 2020.

[7] C. F. Fronczek and J.-Y. Yoon, “Biosensors for monitoring airborne pathogens,” *Journal of laboratory automation*, vol. 20, no. 4, pp. 390–410, 2015.

[8] P. L. Kinney, “Interactions of climate change, air pollution, and human health,” *Current environmental health reports*, vol. 5, pp. 179–186, 2018.

[9] P. Bahl, C. Doolan, C. De Silva, A. A. Chughtai, L. Bourouiba, and C. R. MacIntyre, “Airborne or droplet precautions for health workers treating coronavirus disease 2019?” *The Journal of infectious diseases*, vol. 225, no. 9, pp. 1561–1568, 2022.

[10] S. Herfst, E. J. Schrauwen, M. Linster, S. Chutinimitkul, E. De Wit, V. J. Munster, E. M. Sorrell, T. M. Bestebroer, D. F. Burke, D. J. Smith *et al.*, “Airborne transmission of influenza a/h5n1 virus between ferrets,” *science*, vol. 336, no. 6088, pp. 1534–1541, 2012.

[11] L. Simonsen, P. Spreeuwenberg, R. Lustig, R. J. Taylor, D. M. Fleming, M. Kroneman, M. D. Van Kerkhove, A. W. Mounts, W. J. Paget, and G. C. Teams, “Global mortality estimates for the 2009 influenza pandemic from the glamor project: a modeling study,” *PLoS medicine*, vol. 10, no. 11, p. e1001558, 2013.

[12] E. Rees, V. Ng, P. Gachon, A. Mawudeku, D. McKenney, J. Pedlar, D. Yemshanov, J. Parmely, and J. Knox, “Early detection and prediction of infectious disease outbreaks,” *CCDR*, vol. 45, no. 5, 2019.

[13] N. Peiffer-Smadja, T. M. Rawson, R. Ahmad, A. Buchard, P. Georgiou, F.-X. Lescure, G. Birgand, and A. H. Holmes, “Machine learning for clinical decision support in infectious diseases: a narrative review of current applications,” *Clinical Microbiology and Infection*, vol. 26, no. 5, pp. 584–595, 2020.

[14] L. Morawska, J. W. Tang, W. Bahnfleth, P. M. Bluyssen, A. Boerstra, G. Buonanno, J. Cao, S. Dancer, A. Floto, F. Franchimon *et al.*, "How can airborne transmission of covid-19 indoors be minimised?" *Environment international*, vol. 142, p. 105832, 2020.

[15] R. Vinarti and L. Hederman, "A knowledge-base for a personalized infectious disease risk prediction system," in *Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth.* IOS Press, 2018, pp. 531–535.

[16] E. J. Chow, J. D. Doyle, and T. M. Uyeki, "Influenza virus-related critical illness: prevention, diagnosis, treatment," *Critical Care*, vol. 23, pp. 1–11, 2019.

[17] https://www.hindustantimes.com/lifestyle/health/covid-what-we-know-about-the-omicron-variant-101638246422033.html, November 2021.

[18] D. E. Bloom, S. Black, and R. Rappuoli, "Emerging infectious diseases: A proactive approach," *Proceedings of the National Academy of Sciences*, vol. 114, no. 16, pp. 4055–4059, 2017.

[19] M. Bhatia, S. Kaur, and S. K. Sood, "Iot-inspired smart toilet system for home-based urine infection prediction," *ACM Transactions on Computing for Healthcare*, vol. 1, no. 3, pp. 1–25, 2020.

[20] S. Kumari, M. Bhatia, and G. Stea, "Fog-computing based healthcare framework for predicting encephalitis outbreak," *Big Data Research*, vol. 29, p. 100330, 2022.

[21] A. E. Gorbalenya, S. C. Baker, R. S. Baric, R. J. de Groot, C. Drosten, A. A. Gulyaeva, B. L. Haagmans, C. Lauber, A. M. Leontovich, B. W. Neuman *et al.*, "Severe acute respiratory syndrome-related coronavirus: The species and its viruses–a statement of the coronavirus study group," *BioRxiv*, 2020.

[22] E. National Academies of Sciences, Medicine *et al.*, "Rapid expert consultation on the possibility of bioaerosol spread of sars-cov-2 for the covid-19 pandemic (april 1, 2020)," 2020.

[23] K. T. L. Sy, N. J. L. Haw, and J. Uy, "Previous and active tuberculosis increases risk of death and prolongs recovery in patients with covid-19," *Infectious Diseases*, vol. 52, no. 12, pp. 902–907, 2020.

[24] O. E. Amani Yahiaoui and N. Yumusak, "A new method of automatic recognition for tuberculosis disease diagnosis using support vector machines," *Biomed Res*, vol. 28, pp. 4208–12, 2017.

[25] D. Visca, C. Ong, S. Tiberi, R. Centis, L. D'ambrosio, B. Chen, J. Mueller, P. Mueller, R. Duarte, M. Dalcolmo *et al.*, "Tuberculosis and covid-19 interaction: a review of biological, clinical and public health effects," *Pulmonology*, vol. 27, no. 2, pp. 151–165, 2021.

[26] M. J. Griffiths, L. Turtle, and T. Solomon, "Japanese encephalitis virus infection," *Handbook of clinical neurology*, vol. 123, pp. 561–576, 2014.

[27] L. Ladhani, G. Pardon, H. Meeuws, L. van Wesenbeeck, K. Schmidt, L. Stuyver, and W. van der Wijngaart, "Sampling and detection of airborne influenza virus towards point-of-care applications," *PloS one*, vol. 12, no. 3, p. e0174314, 2017.

[28] K. K. Holmes, S. Bertozzi, B. R. Bloom, P. Jha, H. Gelband, L. M. DeMaria, and S. Horton, "Major infectious diseases: key messages from disease control priorities," *Major infectious diseases*, 2017.

[29] M. D. Decker and K. M. Edwards, "Pertussis (whooping cough)," *The Journal of Infectious Diseases*, vol. 224, no. Supplement_4, pp. S310–S320, 2021.

[30] M. Kupilik and F. Witmer, "Spatio-temporal violent event prediction using gaussian process regression," *Journal of Computational Social Science*, vol. 1, no. 2, pp. 437–451, 2018.

[31] I. Iswanto, E. Wahyudi Setiawan, K. Shankar, and P. T. Nguyen, “Machine learning for healthcare.”

[32] N. Emanet, H. R. Öz, N. Bayram, and D. Delen, “A comparative analysis of machine learning methods for classification type decision problems in healthcare,” *Decision Analytics*, vol. 1, pp. 1–20, 2014.

[33] B. Mahesh, “Machine learning algorithms-a review,” *International Journal of Science and Research (IJSR).[Internet]*, vol. 9, no. 1, pp. 381–386, 2020.

[34] M. Méndez, M. G. Merayo, and M. Núñez, “Machine learning algorithms to forecast air quality: a survey,” *Artificial Intelligence Review*, vol. 56, no. 9, pp. 10 031–10 066, 2023.

[35] R. Sujath, J. M. Chatterjee, and A. E. Hassanien, “A machine learning forecasting model for covid-19 pandemic in india,” *Stochastic Environmental Research and Risk Assessment*, vol. 34, pp. 959–972, 2020.

[36] A. Hernández, A. Panizo, and D. Camacho, “An ensemble algorithm based on deep learning for tuberculosis classification,” in *Intelligent Data Engineering and Automated Learning–IDEAL 2019: 20th International Conference, Manchester, UK, November 14–16, 2019, Proceedings, Part I 20.* Springer, 2019, pp. 145–154.

[37] H. Mukhtar, S. Rubaiee, M. Krichen, and R. Alroobaea, “An iot framework for screening of covid-19 using real-time data from wearable sensors,” *International journal of environmental research and public health*, vol. 18, no. 8, p. 4022, 2021.

[38] P. Singh, N. Singh, K. K. Singh, and A. Singh, “Diagnosing of disease using machine learning,” in *Machine learning and the internet of medical things in healthcare.* Elsevier, 2021, pp. 89–111.

[39] A. S. Poonia, C. Banerjee, A. Banerjee, and S. Sharma, "Aligning misuse case oriented quality requirements metrics with machine learning approach," in *Soft Computing: Theories and Applications: Proceedings of SoCTA 2017*. Springer, 2019, pp. 687–692.

[40] D. M. Abdullah and A. M. Abdulazeez, "Machine learning applications based on svm classification a review," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 81–90, 2021.

[41] P. Palimkar, R. N. Shaw, and A. Ghosh, "Machine learning technique to prognosis diabetes disease: Random forest classifier approach," in *Advanced Computing and Intelligent Technologies: Proceedings of ICACIT 2021*. Springer, 2022, pp. 219–244.

[42] A. M. Posonia, S. Vigneshwari, and D. J. Rani, "Machine learning based diabetes prediction using decision tree j48," in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*. IEEE, 2020, pp. 498–502.

[43] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of k-nearest neighbour (knn) algorithm and its different variants for disease prediction," *Scientific Reports*, vol. 12, no. 1, p. 6256, 2022.

[44] N. Kumar, N. N. Das, D. Gupta, K. Gupta, and J. Bindra, "Efficient automated disease diagnosis using machine learning models," *Journal of healthcare engineering*, vol. 2021, 2021.

[45] Z. Shi and L. S. Pun-Cheng, "Spatiotemporal data clustering: A survey of methods," *ISPRS international journal of geo-information*, vol. 8, no. 3, p. 112, 2019.

[46] P. Haddawy, A. I. Hasan, R. Kasantikul, S. Lawpoolsri, P. Sa-angchai, J. Kaewkungwal, and P. Singhasivanon, "Spatiotemporal bayesian networks

for malaria prediction," *Artificial intelligence in medicine*, vol. 84, pp. 127–138, 2018.

[47] M. R. Nalluri, K. Kannan, M. Manisha, and D. S. Roy, "Hybrid disease diagnosis using multiobjective optimization with evolutionary parameter optimization," *Journal of healthcare engineering*, vol. 2017, 2017.

[48] C.-Y. Ma, M.-H. Chen, Z. Kira, and G. AlRegib, "Ts-lstm and temporal-inception: Exploiting spatiotemporal dynamics for activity recognition," *Signal Processing: Image Communication*, vol. 71, pp. 76–87, 2019.

[49] P. M. Kumar and U. D. Gandhi, "A novel three-tier internet of things architecture with machine learning algorithm for early detection of heart diseases," *Computers & Electrical Engineering*, vol. 65, pp. 222–235, 2018.

[50] K.-H. Kim, E. Kabir, and S. A. Jahan, "Airborne bioaerosols and their impact on human health," *Journal of Environmental sciences*, vol. 67, pp. 23–35, 2018.

[51] S. Sareen, S. K. Sood, and S. K. Gupta, "Iot-based cloud framework to control ebola virus outbreak," *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, pp. 459–476, 2018.

[52] S. K. Sood and I. Mahajan, "Wearable iot sensor based healthcare system for identifying and controlling chikungunya virus," *Computers in Industry*, vol. 91, pp. 33–44, 2017.

[53] N. H. Hassan, E. Salwana, S. M. Drus, N. Maarop, G. N. Samy, and N. A. Ahmad, "Proposed conceptual iot-based patient monitoring sensor for predicting and controlling dengue," *International Journal of Grid and Distributed Computing*, vol. 11, no. 4, pp. 127–134, 2018.

[54] P. James, A. Kunoor, and P. Rakesh, "Awareness of health care workers, patients and visitors regarding air borne infection control–a descriptive study

from a tertiary care centre in kerala, southern india," *indian journal of tuberculosis*, vol. 65, no. 2, pp. 168–171, 2018.

[55] U. S. Shanthamallu, A. Spanias, C. Tepedelenlioglu, and M. Stanley, "A brief survey of machine learning methods and their sensor and iot applications," in *2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA)*. IEEE, 2017, pp. 1–8.

[56] S. Sareen, S. K. Sood, and S. K. Gupta, "An automatic prediction of epileptic seizures using cloud computing and wireless sensor networks," *Journal of medical systems*, vol. 40, pp. 1–18, 2016.

[57] S. Tuli, N. Basumatary, S. S. Gill, M. Kahani, R. C. Arya, G. S. Wander, and R. Buyya, "Healthfog: An ensemble deep learning based smart healthcare system for automatic diagnosis of heart diseases in integrated iot and fog computing environments," *Future Generation Computer Systems*, vol. 104, pp. 187–200, 2020.

[58] R. Tellier, Y. Li, B. J. Cowling, and J. W. Tang, "Recognition of aerosol transmission of infectious agents: a commentary," *BMC infectious diseases*, vol. 19, pp. 1–9, 2019.

[59] W. M. Sweileh, "Bibliometric analysis of peer-reviewed literature on climate change and human health with an emphasis on infectious diseases," *Globalization and health*, vol. 16, no. 1, p. 44, 2020.

[60] H. K. Bharadwaj, A. Agarwal, V. Chamola, N. R. Lakkaniga, V. Hassija, M. Guizani, and B. Sikdar, "A review on the role of machine learning in enabling iot based healthcare applications," *IEEE Access*, vol. 9, pp. 38 859–38 890, 2021.

[61] S. K. Sood and I. Mahajan, "Fog-cloud based cyber-physical system for distinguishing, detecting and preventing mosquito borne diseases," *Future Generation Computer Systems*, vol. 88, pp. 764–775, 2018.

[62] S. Sukhralia, M. Verma, S. Gopirajan, P. Dhanaraj, R. Lal, N. Mehla, and C. R. Kant, "From dengue to zika: the wide spread of mosquito-borne arboviruses," *European Journal of Clinical Microbiology & Infectious Diseases*, vol. 38, pp. 3–14, 2019.

[63] P. Verma and S. K. Sood, "Cloud-centric iot based disease diagnosis healthcare framework," *Journal of Parallel and Distributed Computing*, vol. 116, pp. 27–38, 2018.

[64] P. Martinez, M. Al-Hussein, and R. Ahmad, "A scientometric analysis and critical review of computer vision applications for construction," *Automation in Construction*, vol. 107, p. 102947, 2019.

[65] P. Kokol, H. Blažun Vošner, and J. Završnik, "Application of bibliometrics in medicine: a historical bibliometrics analysis," *Health Information & Libraries Journal*, vol. 38, no. 2, pp. 125–138, 2021.

[66] A. Kumar and G. Kushwaha, "Humanitarian logistics: A review and scientometric analysis," *Journal of Information Technology Research (JITR)*, vol. 11, no. 4, pp. 53–71, 2018.

[67] K. S. Rawat and S. K. Sood, "Emerging trends and global scope of big data analytics: a scientometric analysis," *Quality & Quantity*, vol. 55, no. 4, pp. 1371–1396, 2021.

[68] M. Wang, P. Liu, R. Zhang, Z. Li, and X. Li, "A scientometric analysis of global health research," *International journal of environmental research and public health*, vol. 17, no. 8, p. 2963, 2020.

[69] Z. Noorimotlagh, N. Jaafarzadeh, S. S. Martínez, and S. A. Mirzaee, "A systematic review of possible airborne transmission of the covid-19 virus (sars-cov-2) in the indoor air environment," *Environmental research*, vol. 193, p. 110612, 2021.

[70] J. Song, H. Zhang, and W. Dong, "A review of emerging trends in global ppp research: Analysis and visualization," *Scientometrics*, vol. 107, pp. 1111–1147, 2016.

[71] N. J. Van Eck and L. Waltman, "Citnetexplorer: A new software tool for analyzing and visualizing citation networks," *Journal of informetrics*, vol. 8, no. 4, pp. 802–823, 2014.

[72] V. Rajagopal, S. P. Venkatesan, and M. Goh, "Decision-making models for supply chain risk mitigation: A review," *Computers & Industrial Engineering*, vol. 113, pp. 646–682, 2017.

[73] S. Kumari and M. Bhatia, "A systematic analysis on airborne infectious virus diseases: A review," in *International Conference on Innovations in Data Analytics.* Springer, 2022, pp. 489–498.

[74] J. A. Moral-Muñoz, E. Herrera-Viedma, A. Santisteban-Espejo, and M. J. Cobo, "Software tools for conducting bibliometric analysis in science: An up-to-date review," *Profesional de la información/Information Professional*, vol. 29, no. 1, 2020.

[75] S. Billore and T. Anisimova, "Panic buying research: A systematic literature review and future research agenda," *International Journal of Consumer Studies*, vol. 45, no. 4, pp. 777–804, 2021.

[76] C. Paydas Turan, "Success drivers of co-branding: A meta-analysis," *International Journal of Consumer Studies*, vol. 45, no. 4, pp. 911–936, 2021.

[77] G. Ruggeri, L. Orsi, and S. Corsi, "A bibliometric analysis of the scientific literature on fairtrade labelling," *International Journal of Consumer Studies*, vol. 43, no. 2, pp. 134–152, 2019.

[78] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal of research and development*, vol. 3, no. 3, pp. 210–229, 1959.

[79] S. Kumari and M. Bhatia, “Machine learning techniques for public health system: A scientometric review,” in *2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA).* IEEE, 2022, pp. 1–6.

[80] X. Liu, L. Faes, A. U. Kale, S. K. Wagner, D. J. Fu, A. Bruynseels, T. Mahendiran, G. Moraes, M. Shamdas, C. Kern *et al.*, “A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis,” *The lancet digital health*, vol. 1, no. 6, pp. e271–e297, 2019.

[81] S. R. Nayak, D. R. Nayak, U. Sinha, V. Arora, and R. B. Pachori, “Application of deep learning techniques for detection of covid-19 cases using chest x-ray images: A comprehensive study,” *Biomedical Signal Processing and Control*, vol. 64, p. 102365, 2021.

[82] R. Jain, M. Gupta, S. Taneja, and D. J. Hemanth, “Deep learning based detection and analysis of covid-19 on chest x-ray images,” *Applied Intelligence*, vol. 51, pp. 1690–1700, 2021.

[83] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, “Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images,” *Pattern Recognition Letters*, vol. 138, pp. 638–643, 2020.

[84] I. D. Apostolopoulos and T. A. Mpesiana, “Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks,” *Physical and engineering sciences in medicine*, vol. 43, pp. 635–640, 2020.

[85] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, “Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images,” *arXiv preprint arXiv:2003.11055*, 2020.

[86] H.-t. Zhang, J.-s. Zhang, H.-h. Zhang, Y.-d. Nan, Y. Zhao, E.-q. Fu, Y.-h. Xie, W. Liu, W.-p. Li, H.-j. Zhang *et al.*, "Automated detection and quantification of covid-19 pneumonia: Ct imaging analysis by a deep learning-based software," *European journal of nuclear medicine and molecular imaging*, vol. 47, pp. 2525–2532, 2020.

[87] V. Ramalingam, A. Dandapath, and M. K. Raja, "Heart disease prediction using machine learning techniques: a survey," *International Journal of Engineering & Technology*, vol. 7, no. 2.8, pp. 684–687, 2018.

[88] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Q. Ni, Y. Chen, J. Su *et al.*, "A deep learning system to screen novel coronavirus disease 2019 pneumonia," *Engineering*, vol. 6, no. 10, pp. 1122–1129, 2020.

[89] M. Farooq and A. Hafeez, "Covid-resnet: A deep learning framework for screening of covid19 from radiographs," *arXiv preprint arXiv:2003.14395*, 2020.

[90] S. K. Sood, K. S. Rawat, and D. Kumar, "Scientometric analysis of ict-assisted intelligent control systems response to covid-19 pandemic," *Neural Computing and Applications*, vol. 35, no. 26, pp. 18 829–18 849, 2023.

[91] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," *Pattern Analysis and Applications*, vol. 24, pp. 1207–1220, 2021.

[92] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia, "Covid-19 screening on chest x-ray images using deep learning based anomaly detection," *arXiv preprint arXiv:2003.12338*, vol. 27, no. 10.48550, 2020.

[93] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of covid-19 cases using deep neural networks with x-ray images," *Computers in biology and medicine*, vol. 121, p. 103792, 2020.

[94] M. M. Islam, F. Karray, R. Alhajj, and J. Zeng, "A review on deep learning techniques for the diagnosis of novel coronavirus (covid-19)," *Ieee Access*, vol. 9, pp. 30 551–30 572, 2021.

[95] W. Linda, "A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images," *J. Netw. Comput. Appl*, vol. 20, pp. 1–12, 2020.

[96] C. Liu, Y. Cao, M. Alcantara, B. Liu, M. Brunette, J. Peinado, and W. Curioso, "Tx-cnn: Detecting tuberculosis in chest x-ray images using convolutional neural network," in *2017 IEEE international conference on image processing (ICIP)*. IEEE, 2017, pp. 2314–2318.

[97] E. Luz, P. Silva, R. Silva, L. Silva, J. Guimarães, G. Miozzo, G. Moreira, and D. Menotti, "Towards an effective and efficient deep learning model for covid-19 patterns detection in x-ray images," *Research on Biomedical Engineering*, pp. 1–14, 2021.

[98] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, R. Wang, H. Zhao, Y. Chong *et al.*, "Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 18, no. 6, pp. 2775–2780, 2021.

[99] L. Li, L. Qin, Z. Xu, Y. Yin, and X. Wang, "Bin kong, junjie bai, yi lu, zhenghan fang, qi song, et al," *Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. Radiology*, vol. 5, 2020.

[100] F. Shi, L. Xia, F. Shan, B. Song, D. Wu, Y. Wei, H. Yuan, H. Jiang, Y. He, Y. Gao *et al.*, "Large-scale screening to distinguish between covid-19 and community-acquired pneumonia using infection size-aware classification," *Physics in medicine & Biology*, vol. 66, no. 6, p. 065031, 2021.

[101] S. S. Meraj, R. Yaakob, A. Azman, S. N. M. Rum, and A. A. Nazri, "Artificial intelligence in diagnosing tuberculosis: a review," *International Journal*

on Advanced Science, Engineering and Information Technology, vol. 9, no. 1, pp. 81–91, 2019.

[102] Z. Tang, W. Zhao, X. Xie, Z. Zhong, F. Shi, T. Ma, J. Liu, and D. Shen, "Severity assessment of covid-19 using ct image features and laboratory indices," *Physics in Medicine & Biology*, vol. 66, no. 3, p. 035015, 2021.

[103] Y.-H. Chan, Y.-Z. Zeng, H.-C. Wu, M.-C. Wu, H.-M. Sun *et al.*, "Effective pneumothorax detection for chest x-ray images using local binary pattern and support vector machine," *Journal of healthcare engineering*, vol. 2018, 2018.

[104] J.-H. Huang, C.-H. H. Yang, F. Liu, M. Tian, Y.-C. Liu, T.-W. Wu, I. Lin, K. Wang, H. Morikawa, H. Chang *et al.*, "Deepopht: medical report generation for retinal images via deep models and visual explanation," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2021, pp. 2442–2452.

[105] E. Pahwa, D. Mehta, S. Kapadia, D. Jain, and A. Luthra, "Medskip: Medical report generation using skip connections and integrated attention," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3409–3415.

[106] L. Bai, D. Yang, X. Wang, L. Tong, X. Zhu, N. Zhong, C. Bai, C. A. Powell, R. Chen, J. Zhou *et al.*, "Chinese experts' consensus on the internet of things-aided diagnosis and treatment of coronavirus disease 2019 (covid-19)," *Clinical eHealth*, vol. 3, pp. 7–15, 2020.

[107] O. Yildirim, M. Talo, B. Ay, U. B. Baloglu, G. Aydin, and U. R. Acharya, "Automated detection of diabetic subject using pre-trained 2d-cnn models with frequency spectrum images extracted from heart rate signals," *Computers in biology and medicine*, vol. 113, p. 103387, 2019.

[108] U.-O. Dorj, K.-K. Lee, J.-Y. Choi, and M. Lee, "The skin cancer classification using deep convolutional neural network," *Multimedia Tools and Applications*, vol. 77, pp. 9909–9924, 2018.

[109] A. Brunetti, L. Carnimeo, G. F. Trotta, and V. Bevilacqua, "Computer-assisted frameworks for classification of liver, breast and blood neoplasias via neural networks: A survey based on medical images," *Neurocomputing*, vol. 335, pp. 274–298, 2019.

[110] R. Zhou, X. Chang, L. Shi, Y.-D. Shen, Y. Yang, and F. Nie, "Person reidentification via multi-feature fusion with adaptive graph learning," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 5, pp. 1592–1601, 2019.

[111] A. M. Alqudah, S. Qazan, and A. Alqudah, "Automated systems for detection of covid-19 using chest x-ray images and lightweight convolutional neural networks," 2020.

[112] F. Ucar and D. Korkmaz, "Covidiagnosis-net: Deep bayes-squeezenet based diagnosis of the coronavirus disease 2019 (covid-19) from x-ray images," *Medical hypotheses*, vol. 140, p. 109761, 2020.

[113] Z. Li, F. Nie, X. Chang, L. Nie, H. Zhang, and Y. Yang, "Rank-constrained spectral clustering with flexible embedding," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 12, pp. 6073–6082, 2018.

[114] X. Chang, F. Nie, S. Wang, Y. Yang, X. Zhou, and C. Zhang, "Compound rank-k projections for bilinear analysis," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 7, pp. 1502–1513, 2015.

[115] M. Luo, X. Chang, L. Nie, Y. Yang, A. G. Hauptmann, and Q. Zheng, "An adaptive semisupervised feature analysis for video semantic recognition," *IEEE transactions on cybernetics*, vol. 48, no. 2, pp. 648–660, 2017.

[116] P. Ren, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, X. Chen, and X. Wang, "A comprehensive survey of neural architecture search: Challenges and solutions," *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1–34, 2021.

[117] E. Yu, J. Sun, J. Li, X. Chang, X.-H. Han, and A. G. Hauptmann, "Adaptive semi-supervised feature selection for cross-modal retrieval," *IEEE Transactions on Multimedia*, vol. 21, no. 5, pp. 1276–1288, 2018.

[118] Z. Li, L. Yao, X. Chang, K. Zhan, J. Sun, and H. Zhang, "Zero-shot event detection via event-adaptive concept relevance mining," *Pattern Recognition*, vol. 88, pp. 595–603, 2019.

[119] H. Panwar, P. Gupta, M. K. Siddiqui, R. Morales-Menendez, and V. Singh, "Application of deep learning for fast detection of covid-19 in x-rays using ncovnet," *Chaos, Solitons & Fractals*, vol. 138, p. 109944, 2020.

[120] P. K. Sethy and S. K. Behera, "Detection of coronavirus disease (covid-19) based on deep features," 2020.

[121] T. Mahmud, M. A. Rahman, and S. A. Fattah, "Covxnet: A multi-dilation convolutional neural network for automatic covid-19 and other pneumonia detection from chest x-ray images with transferable multi-receptive feature optimization," *Computers in biology and medicine*, vol. 122, p. 103869, 2020.

[122] M. Loey, F. Smarandache, and N. E. M. Khalifa, "Within the lack of chest covid-19 x-ray dataset: a novel detection model based on gan and deep transfer learning," *Symmetry*, vol. 12, no. 4, p. 651, 2020.

[123] S. R. Abdani, M. A. Zulkifley, and N. H. Zulkifley, "A lightweight deep learning model for covid-19 detection," in *2020 IEEE Symposium on Industrial Electronics & Applications (ISIEA)*. IEEE, 2020, pp. 1–5.

[124] A. A. Aliabadi, S. N. Rogak, K. H. Bartlett, and S. I. Green, "Preventing airborne disease transmission: review of methods for ventilation design in health care facilities," *Advances in preventive medicine*, vol. 2011, 2011.

[125] C. F. Dillon and M. B. Dillon, “Multiscale airborne infectious disease transmission,” *Applied and Environmental Microbiology*, vol. 87, no. 4, pp. e02 314–20, 2021.

[126] C. Abi Nader, N. Ayache, P. Robert, M. Lorenzi, A. D. N. Initiative *et al.*, “Monotonic gaussian process for spatio-temporal disease progression modeling in brain imaging data,” *Neuroimage*, vol. 205, p. 116266, 2020.

[127] R. Asadi and A. Regan, “A convolution recurrent autoencoder for spatio-temporal missing data imputation,” *arXiv preprint arXiv:1904.12413*, 2019.

[128] Z. Li, F. Nie, X. Chang, Y. Yang, C. Zhang, and N. Sebe, “Dynamic affinity graph construction for spectral clustering using multiple features,” *IEEE transactions on neural networks and learning systems*, vol. 29, no. 12, pp. 6323–6332, 2018.

[129] S. Ahmed, M. F. Hossain, and M. B. T. Noor, “Convid-net: an enhanced convolutional neural network framework for covid-19 detection from x-ray images,” in *Proceedings of International Conference on Trends in Computational and Cognitive Engineering: Proceedings of TCCE 2020.* Springer, 2021, pp. 671–681.

[130] C. Bailer, T. Habtegebrial, D. Stricker *et al.*, “Fast feature extraction with cnns with pooling layers,” *arXiv preprint arXiv:1805.03096*, 2018.

[131] Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, S. Ahmed, and J. Lu, “Brain tumor classification for mr images using transfer learning and fine-tuning,” *Computerized Medical Imaging and Graphics*, vol. 75, pp. 34–46, 2019.

[132] V. Chouhan, S. K. Singh, A. Khamparia, D. Gupta, P. Tiwari, C. Moreira, R. Damaševičius, and V. H. C. De Albuquerque, “A novel transfer learning based approach for pneumonia detection in chest x-ray images,” *Applied Sciences*, vol. 10, no. 2, p. 559, 2020.

[133] S. Kumari and M. Bhatia, “A cognitive framework based on deep neural network for classification of coronavirus disease,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 9, pp. 12 075–12 089, 2023.

[134] J. P. Cohen, P. Morrison, and L. Dao, “Covid-19 image data collection,” *arXiv preprint arXiv:2003.11597*, 2020.

[135] X. Li, C. Li, and D. Zhu, “Covid-mobilexpert: On-device covid-19 screening using snapshots of chest x-ray,” 2020.

[136] L. Wang, Z. Q. Lin, and A. Wong, “Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images,” *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.

[137] W. Hariri and A. Narin, “Deep neural networks for covid-19 detection and diagnosis using images and acoustic-based techniques: a recent review,” *Soft computing*, vol. 25, no. 24, pp. 15 345–15 362, 2021.

[138] N. N. Das, N. Kumar, M. Kaur, V. Kumar, and D. Singh, “Automated deep transfer learning-based approach for detection of covid-19 infection in chest x-rays,” *Irbm*, 2020.

[139] K. Verma, G. Sikka, A. Swaraj, S. Kumar, and A. Kumar, “Classification of covid-19 on chest x-ray images using deep learning model with histogram equalization and lung segmentation,” *SN Computer Science*, vol. 5, no. 4, p. 379, 2024.

[140] Z. N. Khan, “Detection of covid19 in chest x-ray images using transfer learning,” *arXiv preprint arXiv:2304.04161*, 2023.

[141] B. Samir, S. Mwanahija, B. Soumia, and U. Özkaya, “Deep learning for classification of chest x-ray images (covid 19),” *arXiv preprint arXiv:2301.02468*, 2023.

[142] A. Kulkarni, G. Parasnis, H. Balasubramanian, V. Jain, A. Chokshi, and R. Sonkusare, "Advancing diagnostic precision: Leveraging machine learning techniques for accurate detection of covid-19, pneumonia, and tuberculosis in chest x-ray images," *arXiv preprint arXiv:2310.06080*, 2023.

[143] I. Mwendo, K. Gikunda, and A. Maina, "Deep transfer learning for detecting covid-19, pneumonia and tuberculosis using cxr images–a review," *arXiv preprint arXiv:2303.16754*, 2023.

[144] S. Manimurugan, "Iot-fog-cloud model for anomaly detection using improved naïve bayes and principal component analysis," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–10, 2021.

[145] N. A. Mansour, A. I. Saleh, M. Badawy, and H. A. Ali, "Accurate detection of covid-19 patients based on feature correlated naïve bayes (fcnb) classification strategy," *Journal of ambient intelligence and humanized computing*, pp. 1–33, 2022.

[146] E. A. Varouchakis, G. A. Corzo, G. P. Karatzas, and A. Kotsopoulou, "Spatio-temporal analysis of annual rainfall in crete, greece," *Acta Geophysica*, vol. 66, no. 3, pp. 319–328, 2018.

[147] R. Ferdousi, M. A. Hossain, and A. El Saddik, "Early-stage risk prediction of non-communicable disease using machine learning in health cps," *IEEE Access*, vol. 9, pp. 96 823–96 837, 2021.

[148] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *Journal of Big data*, vol. 6, no. 1, pp. 1–18, 2019.

[149] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE access*, vol. 7, pp. 81 542–81 554, 2019.

[150] A. Gupta and A. Singh, "Prediction framework on early urine infection in iot–fog environment using xgboost ensemble model," *Wireless Personal Communications*, pp. 1–19, 2023.

[151] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, and M. Kaur, "Classification of the covid-19 infected patients using densenet201 based deep transfer learning," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 15, pp. 5682–5689, 2021.

[152] A. Manocha, G. Kumar, and M. Bhatia, "Iot analytics-inspired real-time monitoring for early prediction of covid-19 symptoms," *The Computer Journal*, vol. 66, no. 1, pp. 144–159, 2023.

[153] R. Hertel and R. Benlamri, "Cov-snet: A deep learning model for x-ray-based covid-19 classification," *Informatics in Medicine Unlocked*, vol. 24, p. 100620, 2021.

[154] J. P. B. Saputra and R. P. Bernarte, "The naive bayes algorithm in predicting the spread of the omicron variant of covid-19 in indonesia: Implementation and analysis," *International Journal of Informatics and Information Systems*, vol. 5, no. 2, pp. 84–91, 2022.

[155] S. Kumari, H. Kaur, and P. Gupta, "A cognitive effective framework for analysis, monitoring and identifying variant of coronavirus disease," *The Journal of Supercomputing*, vol. 80.

PUBLICATIONS

International Conferences

1) Sapna Kumari, Munish Bhatia “Machine Learning Techniques For Public Health System: A Scientometric Review”, 2022 Second International Conference on Computer Science, Engineering and Applications, **Published in IEEE conference**

Paper Published in **IEEE Conference “2nd International Conference on Computer Science Engineering and Applications (ICCSEA-2022)“**

2) Sapna Kumari, Munish Bhatia “A Systematic Analysis on Airborne Infectious Virus Diseases: A Review”. ”1st International Conference on Innovations in Data Analytics (ICIDA 2022)“ [**Published in Springer Conference**]

Paper Published in **Springer Conference** “ **1st International Conference on Innovations in Data Analytics (ICIDA 2022)”**

Journals

1) Sapna Kumari and Munish Bhatia Paper Published in “Journal of Ambient Intelligence and Humanized Computing”, Springer, 2022 entitled “A cognitive framework based on deep neural network for classification of coronavirus disease” **((SCI) [Published]**)

2) Sapna Kumari, Harpreet Kaur, Prabal Gupta entitled “A Cognitive Effective Framework for Analysis, Monitoring and Identifying Variant of Coronavirus”, in ”The Journal of Supercomputing”, Springer **((SCIE) [Published**]

Book Chapter

1) **Sapna Kumari**, Harpreet Kaur, Ramandeep Sandhu, and Deepika Ghai, A Book Chapter Published (Free of cost) in “Innovations in Biomedical Engineering: Trends in Scientific Advances and Applications”, Elsevier “A Cognitive technique for human Monkeypox disease diagnosis and Prevention in Healthcare System” [**Published**]

Patent

1) **Sapna Kumari and Harpreet Kaur** Patent Published (Application no. 202311045609) in 2023 “An IoT and CNN-based early detection System for Monkeypox Disease” [**Published**]