Performance Analysis for the Diagnosis of COVID-19 Prediction by Mathematical Modeling & Simulation

Mehtab Siddiqui[1], Hamid Ali Kalwar[2], Muhammad Zohaib Khan[3], Muhammad Ali Khan[4], Aisha Imroz[5], Muhammad Ahmed Kalwar[6], Hussain Bux Marri[7]
[1]Faculty of Health Sciences, Iqra University, North Campus, Karachi, Pakistan
[2]Murshid Hospital & Health Care Centre, Karachi, Pakistan
[3]Shaheed Mohtarma Benazir Bhutto Institute of Trauma (SMBBIT), Karachi, Pakistan
[4]Mehran University of Engineering & Technology, Jamshoro, Pakistan
[5]Avanza (Pvt.) Ltd, Karachi, Pakistan
[6]Shafi (Pvt.) Limited Company, Lahore, Pakistan
[7]Benazir Bhutto Shaheed University of Technology and Skill Development (BBSUTSD), Khairpur Mirs, Pakistan

Email: mehtabsiddiqui707@gmail.com, hamidalikalwar@gmail.com,zohaibkhanpk2017@gmail.com, muhammad.nagar@faculty.muet.edu.pk, aishaimrozpk@gmail.com, kalwar.muhammad.ahmed@gmail.com hussainbux@bbsutsd.edu.pk

Corresponding Author: Muhammad Ahmed Kalwar, kalwar.muhammad.ahmed@gmail.com

Received: 25-02-2023; **Accepted**: 17-05-2023; **Published**: 01-07-2023

Abstract: Machine learning is a type of artificial intelligence (AI) method. The system can be conscious and adaptable to changes in concepts when emphatically programmed for presence based on prior experience. The study built and examined several sorts of ML algorithms in this work by creating a classifier depending on a predefined model to estimate COVID-19. The ML approach might be used in any sector of medicine, including diagnostics. COVID-19 forecast research is still needed, and there is a need to improve the diagnostic COVID-19 methodology. Several investigations have varied objectives regarding COVID-19 estimation methods that employ techniques. The purpose of this effort was to provide an integrative strategy to enhance the accuracy, precision, and recall of the COVID-19 proposed method. This study supports the utilization of ML and DL approaches. It is an essential element in the analysis of ML approaches such as K-Means and is applied to both supervised and unsupervised learning concepts. Several of these ML algorithms use various approaches to examine and perform the Decision Tree ,Naive Bayes and Multi-Layer Perceptron algorithms. Equivalent data sets were examined for clustering and classification approaches, including K-Means techniques. It was demonstrated that outcomes could be foreseen that were efficient, effective, and accurate.The best algorithmic outcome for COVID-19 case detection was achieved by combining K-Means with Multi-Layer Perceptrons (MLPs), resulting in an accuracy of 99.88%.

Keywords: Machine Learning (ML) Deep Learning (DL), K-Means,Multi-layer perceptron (MLP), Decision Tree (DT), Naïve Bayes (NB) Algorithms.

1. Introduction

The COVID-19 disease outbreak, also known as the deadly virus, seems to be a widespread phenomenon of corona diseases in 2019 caused with (SARS-CoV-2). The unique virus was transmitted throughout an incidence in China, in Dec 2019. Measures to restrict it were inadequate, permitting the disease to spread globally. Its Health Organization proclaimed a Critical Serious Infection of Worldwide Significance upon January 30, 2020, and even a virus in March 11, 2020. As of May 6, 2022, the virus has caused over 516 shedload injuries and 6.24 shedload premature deaths, considering it among the deadliest in existence. COVID-19 indications varied as non-existent to lethal, with temperature, chest infection, and tiredness being the most common symptoms. People over the age of 65, and others with chronic health issues, are at a greater risk of developing a serious condition. COVID-19 originates because individuals inhaled unclean environment comprising virus-carrying globules and tiny hazardous air pollutants. Even while they can be absorbed in larger locations, such as underground, the risk of inhaling them is greatest when individuals are seated close together. If infectious fluids enter the eye, nose, or mouth, or, in rare cases, if infectious material comes into contact with the eye, nose, or mouth, reproduction can occur. People who have the sickness remain contagious for 10 days, which means they can spread germs even if they are asymptomatic. Changes have resulted in a slew of bacterial species (varieties) with varying infection rates and morbidity rates [1]. Recently, a decision has been made regarding vaccinated individuals traveling within the United States. According to this decision, those who have been vaccinated against a particular infectious disease will no longer be required to undergo quarantine or testing before or after arriving at their destination within the country. This means that vaccinated individuals can freely travel without these restrictions, assuming they meet the vaccination criteria. It is important to note that this information is based on the latest available knowledge up until September 2021, and there may have been updates or changes in guidelines since then [2].

The COVID-19 epidemic has resulted in a significant number of cases, hospitalizations, and deaths in the US. Between January 1, 2020, and February 5, 2022, there have been over 76.4 million reported cases, 4.4 million new hospitalizations, and 900,000 deaths related to COVID-19. In December 2021, the US Food and Drug Administration (FDA) approved two oral antiviral therapies (OAV) for individuals aged 12 and above who have mild or moderate disease. These therapies are intended for outpatient treatment of early COVID-19 within 5 days of symptom onset, with the aim of preventing the progression to severe disease [4]. This epidemic continues to pressure health devices throughout the global in a multitude of situations, involving rapid growth for emergency rooms and acute limitations of diagnostic instruments, as well as the illness by several clinical staff. As a response, the capability to make appropriate treatment decisions and efficiently employ health services is necessary. RT-PCR, the far more accurate COVID-19 test, has often remained unavailable in impoverished nations [5]–[7]. This contributes to greater disease incidence and delays critical prevention measures. Evaluation capabilities allow the early and precise detection of COVID-19, decreasing the burden on health systems. Monitoring and forecasting that combine multiple signs to evaluate the likelihood of disease have indeed been developed to assist physicians worldwide in categorizing cases, especially throughout situations involving restricted health services. Computerized radiography scanning, somatic manifestations, scientific examine, and a combination of such parameters are used in these simulations.Nonetheless, since the majority of older models were data is collected from hospital - acquired infections, they are useless for detecting for (SARS-CoV-2) inside the entire population [8]–[10].

2. Literature Review

Numerous studies and investigations regarding the examination and identification of COVID-19 have been conducted recently. The initial portion of this section explores the difficulties in applying deep learning techniques to CT scans and chest X-ray pictures for COVID-19 identification [64]. The second section examines pertinent literature to evaluate forecasts for the number of cases that were confirmed, recoveries, and fatalities linked to COVID-19 in the future. COVID-19 has turned into a pandemic as a result of its quick worldwide dissemination. People infected with the virus may not immediately show any symptoms, making detection difficult [22], [65]–[67]. Therefore, in order to put suitable precautions in place, a technique for periodically assessing the number of possibly affected people must be developed [68], [69]. The potential of AI to be included in a model that has been trained for the categorization of hidden pictures gives a considerable benefit. In this work, AI was used to examine a patient's chest X-ray picture and identify whether they are COVID-19 positive. Over two hundred countries throughout the world have been required to respond quickly and thoroughly to the COVID-19 pandemic produced by the virus known as SARS-CoV-2. Between 2 and 20 million people were infected during the first four-month period of the pandemic, which led to at least two hundred thousand fatalities. Governments throughout the world adopted strict regulations, notably the isolation of millions of residents, to slow the virus' fast spread among humans [70]–[72]. However, it is difficult to distinguish between those who have positive and negative COVID-19 instances based alone on the many disease-related symptoms. Despite several efforts, it is difficult to distinguish between the two. Consequently, it is thought to be essential to use tests that can detect the SARS-CoV-2 virus. These tests are essential for detecting cases of the virus that warrant further investigation and aid in the implementation of strategies to stop the virus from spreading [73]–[75]. Chest CT scans, for example, are very helpful and crucial in determining the stage and possible hazards of COVID-19 in patients' lungs [70], [76], [77]. To reduce transmission from person to person and improve patient care, Covid-19 must be diagnosed as soon as possible. The most effective method now available to stop the spreading of COVID-19 is to isolate individuals in good health from those who suffer from or who are suspected of harbouring the virus [63], [76], [78]. Approaches for machine learning role demonstrated crucial comprehensions of the COVID-19 diagnostic, including the lungs computed tomography (CT) scan's potential as either the first screening test or a substitute.

Khan et al. reviewed, analyze & compared the performance of various ML and DL techniques for detecting anomalies in credit card frauds [11]. It is evident by the publications of many other researchers [12]–[16]. In the initial diagnosis, neither absence nor presence of any sign or symptom have been found accurate enough to rule in or rule out disease [17]–[19]. The research is evident that the prediction performance of diagnosis analysis in COVID-19 can be improved by mathematical modelling, simulation and machine learning [20]–[25]. The study suggests that both men and women have an equal likelihood of contracting COVID-19, meaning the prevalence of the virus is similar among both genders. However, it also highlights that men who become infected with COVID-19 are more vulnerable to experiencing severe outcomes and a higher risk of death, regardless of their age. In other words, even when considering age as a factor, men tend to face worse consequences from the disease compared to women [26]–[31]. A multitude of machine-learning systems for artificial intelligence are in current use, particularly in industry. Some need supervised learning, while others are unsupervised; the Artificial Intelligence trend in this age is bringing artificial intelligence to the attention of the globe. A subset of machine learning. Circular and square circles are the two forms of circles. We may feed the machine a variety of data [32]. Learning oprganizations have the proven records of performance improvement in organizational operations by the implementation of automization & optimization techniques [33]–[39].

Zaman et al. compared the comparative study of relational & non relational Database Management Systems (DBMS) and compared their speed, efficiency and performance were monitored [40]. Khan et al. also compared the relational & non relational DBMS i.e. SQL & NoSQL databases. Both were compared in terms of loading, response, and retrieval times to discover their smootheness, efficiency and performance [41]. Khan et al. conducted the study and utilized different AI & ML techniques to assess the performance of three algorithms i.e. K-Means, KNN & SVM. They conducted simulations to demonstrate their performance with results. The study aimed to showcase effective algorithms and their capabilities in terms of performance computation [42]. Many organizations in Pakistan lags behind in the optimization of operations [43] & [44] by applications of AI & ML. The recent applied case studies of Pakistani organizations in the context of optimization by automation include procurement report [45], routine report making [46], purchase order [47], acquisition report [45], planning report [48], Supplier Price Evaluation Report [49], material delivery time analysis [50], product mix & profit maximization [51], and order costing analysis [52]. The ERP system at logistic centre of CAA Karachi is reduced by 50% of working time through the use of VBA [53]. Kalwar & Khan saved Seventy-five percent (75%) of the staff time which would have otherwise spent manually creating the procurement report was saved by automation in operations. The process of generating the order costing report was significantly faster after automation compared to the manual technique, with a time reduction of 8592% [47]. More recently, the automation was implanted at material cost comparative analysis (MCCA) in an industry which resulted in 100% accuracy and 72.20% time reduction in 58.51 minutes [54]. Mathematical modeling & simulation is getting more improtance day by day in hospitals's outpatient departments (OPDs) and emergency sections. Practioners & researchers also developed and applied the queuing models of operations research to optimize the performance in hospitals's outpatient departments (OPDs) and emergency sections [55]–[60]. Kalwar et al. and Khaskheli et al. conducted the comparative study of queuing systems of out patient departments (OPDs) in two large public hospitals of Pakistan. Researchers highlighted the potential applications of Mathematical modeling, simulation & queuing models of operations research to optimize the performance in hospitals's outpatient departments (OPDs) and emergency Health Care Units of Pakistan [61], [62]. [63]

The utilization of real-time inverse transcriptase-polymerase chain reaction (RT-PCR) is crucial in distinguishing COVID-19 pneumonia from other forms of viral pneumonia. Additionally, the examination of lung CT scans plays a significant role in identifying the disparities between these conditions [9], [79], [80]. Healthcare facilities place a high value on clinical task prediction, and several fields are employing computer-aided clinical predictive models, such as determining the likelihood of cardiac failure [81], [82], mortality in pneumonia [83]–[85], mortality risk in critical care [86], [87]. These technologies help medical professionals properly understand and evaluate clinical results. In this research, the authors propose a clinical COVID-19 prediction model by building on earlier methodological advancements. There are not many comparable research studies on COVID-19 prediction in clinical trials in the literature. Machine learning techniques were used in researchers to predict the clinical impact of coronavirus [88], [89]. Cangnan People's Hospital in Wenzhou and Wenzhou Central Hospital, China provided the data used in the study [90]. However, because the information is private, access to it is limited. The study used 6 distinct classifiers, including, KNN, Logistic Regression, 2different decision trees, SVM, Random Forests and took into account eleven clinical characteristics. Based only on accuracy scores, the classifiers' performance was evaluated. The SVM classifier had an accuracy rate of 80%. In a separate investigation, Researchers forecasted the COVID-19 diagnostic using ML classifiers [91], [92]. Clinical data used in the study was sourced from Hospital Israelita Albert Einstein in Sao Paulo, Brazil [93]. The study took into account 18 clinical observations, and the classifiers were assessed based on metrics such as specificity, sensitivity, AUC, Brier score, negative predictive value, F1-score and positive predictive value. Five distinct classifiers were employed, namely SVM, random forests, neural networks, logistic regression, and gradient boosted trees. The random forest and support Vector Machine classifiers both achieved the highest AUC i.e. 0.847. In another study, a clinical predictive model was introduced for

COVID-19 [94]. Similarly to the study mentioned earlier in which the data was collected from the Hospital Israelita Albert Einstein in Sao Paulo, Brazil [93]. The authors employed multiple machine learning techniques, such as XFB, RF, SVM, NN, and LR. The AUC sensitivity and specificity values of these classifiers were calculated to assess their performance. Having an AUC score of 66%, the XGB classifier outperformed the rest of them [94].

It was found that the severity and mortality was variable in COVID-19 patients with different genders [95]. Effective prediction models have been developed to assist in the initial screening of COVID-19 through the evaluation of basic clinical signs and symptoms. These models aim to improve the prioritization of clinical resources, which can help alleviate the overwhelming strain on healthcare systems. By accurately identifying individuals who are likely to have COVID-19, these prediction models allow for more efficient allocation of testing and treatment resources, ultimately reducing the burden on healthcare providers and facilities [22], [96]–[98]. The elements of statistical learning i.e. data mining, inference & prediction can be used effectively in the diseases prediction like covid 19 [99]–[103]. Mathematical modeling, simulation, machine learning, artificial intelligence and relevant datasets should be used in the propagation analysis, prediction, diagnosis, prognosis and management COVID-19 [23], [24], [104], [105]. Gradient boosting (GB) is a highly effective technique for predicting tabular data and is widely regarded as the state of the art in the field of machine learning. It has been employed by numerous successful algorithms. GB involves combining multiple weak predictive models, such as decision trees, into an ensemble model. Each weak model is trained on the errors or residuals of the previous models, with the aim of minimizing the overall prediction error. This iterative process creates a strong predictive model that can capture complex relationships in the data. GB has proven to be particularly effective in various domains, including regression and classification tasks, and has achieved notable success in competitions and real-world applications [22], [104], [106]–[109]. Multivariate imputation has recently been shown to work in combination with predictive models in different missingness scenarios. In previous studies, it has been suggested that gradient-boosting predictors are effective in handling missing values in various predictions, including those related to COVID-19. This implies that when using a gradient-boosting model to make predictions, it has the ability to handle missing data points without requiring explicit imputation or preprocessing techniques. The model's inherent nature allows it to effectively learn from the available data and make accurate predictions even when some data points are missing [110]–[114]. The assumption commonly made is that data points exist on a manifold, which is a lower-dimensional structure within a higher-dimensional space. Based on this assumption, various studies have focused on density estimation methods specifically tailored to the set of nodes in a graph. In the context of graphs, density estimation refers to estimating the distribution of data points on the graph. However, instead of considering the entire ambient space, the estimation is restricted to the support of the graph nodes. This means that the density estimation techniques aim to capture the underlying distribution of data points within the graph structure, considering the connections and relationships between the nodes. By assuming that the data points lie on a low-dimensional manifold within the graph, researchers have developed approaches to estimate the density of the data within this limited space. These methods take into account the graph topology and utilize techniques such as spectral analysis, graph diffusion, or graph Laplacians to model the density distribution. The objective of these density estimation techniques is to provide insights into the underlying data distribution, identify clusters or patterns within the graph, and enable various graph-based analyses or applications. By focusing on the support of the graph nodes, these methods leverage the inherent structure of the graph to improve the accuracy of density estimation compared to traditional approaches that consider the entire ambient space [66], [115]–[120]. Baig et al. investigated the efficacy of ML techniques in forecasting the success of Higher Education Institutions' (HEIs) students. ML & deep learning algorithms were investigated separately and in combinations in forecasting the success of students in their classes [121]. Channar et al. highlighted knowledge sharing factors in the Higher Education Institutions (HEIs) [122]. Information seeking behaviour is changing dynamically [123], [124] and the information retrieval patterns of the students and faculty members in Higher Education Institutions (HEIs) are also change. Advanced technologies like AI & ML promoted the effective use of Social Media [125] also in Information seeking/retrieval in Higher Education Institutions (HEIs) [126]–[128]. Khan et al. conducted a performance analysis of different ML & DL techniques, including PCA, data mining, Fuzzy, LR, DT, & NB algorithms, to detect anomalies in credit card frauds. The study aimed to compare and evaluate the effectiveness of these techniques and potentially improve the accuracy of fraud detection systems in the financial domain [11].

3. Problem statement

In the field of medical science, there is a growing need to improve the healthcare system, particularly in the context of COVID-19. By the use of the approaches mentioned in detailed literarure review, the authors will examine both unsupervised and supervised learning methods, followed by K-Means, and prediction techniques utilized in MLP, DT, and NB algorithms. These algorithms are utilized in clustering and classification procedures.

The objective is to reduce mortality rates and enhance the efficiency of COVID-19 diagnosis through advanced research and development. To achieve this, it is crucial to explore various machine learning approaches and prediction models that can accurately predict COVID-19 outcomes using the available dataset. The performance of these models needs to be evaluated and measured in terms of f1-score. recall qualityprecision and accuracy.This research aims to address these challenges by utilizing the COVID-19 dataset and analyzing the performance of different ML algorithms in developing prediction models.There are four segments to this paper. The first part provides an introduction to the research study. Second segment discusses the work related to this research. Third part focuses on the results and discussion of various algorithm combinations used for predicting the coronavirus. Lastly, the concluding section presents a statement on the most efficient algorithm combination.

Nowadays, many researchers now work in medical science, with doctors and scientists seeking to better the health-care system. These research domains aid the health-care system because there are less incidences of covid-19 patients dying, and anybody can quickly diagnose covid-19 prediction and increasing diagnostic covid-19 prediction model output quality, accuracy, precisions, recall, and f1-score.Corona Virus Diasease-2019, is a Coronaviridae virus. A non-vaccinable virus is wreaking havoc on people's lives as well as financial and economic institutions all across the world. Everything in society has been placed on hold mercilessly. The COVID-19 dataset was utilised in this work to determine which ML approaches are used to develop various prediction models, and their performance was calculated and evaluated.

4. Research methodology

In this section, the work breakdown structure of the COVID-19 prediction system is briefly defined. The steps involved are as follows:

1. Dataset Retrieval: Initially, the dataset is retrieved from Google Drive, which serves as the source of data for the system.
2. Data Cleaning and Feature Extraction: The retrieved data undergoes cleaning to remove any inconsistencies or irrelevant information. Feature extraction is performed using the TfidfTransformer method, which helps in representing the data in a meaningful way.
3. Pre-processing and Standardization: The data is pre-processed to prepare it for analysis. Additionally, the data is standardized using the MinMaxScaler method, which ensures that all features are transformed to a consistent scale.
4. K-Means Clustering: This technique is an unsupervised ML method, is applied to the pre-processed and standardized data. This method enhances model accuracy, precision, recall, and F1-score.
5. Train and Test Data Split: The dataset is divided into test data and training data to evaluate the performance of the model.
6. Algorithm Implementation: The NB ,MLP and DT algorithms are utilized to create and construct the final prediction model for COVID-19.

Figure 4 illustrates the architecture of the COVID-19 prediction system, providing a visual representation that aids in understanding the research activity and serves as a concise outline of the work breakdown structure.

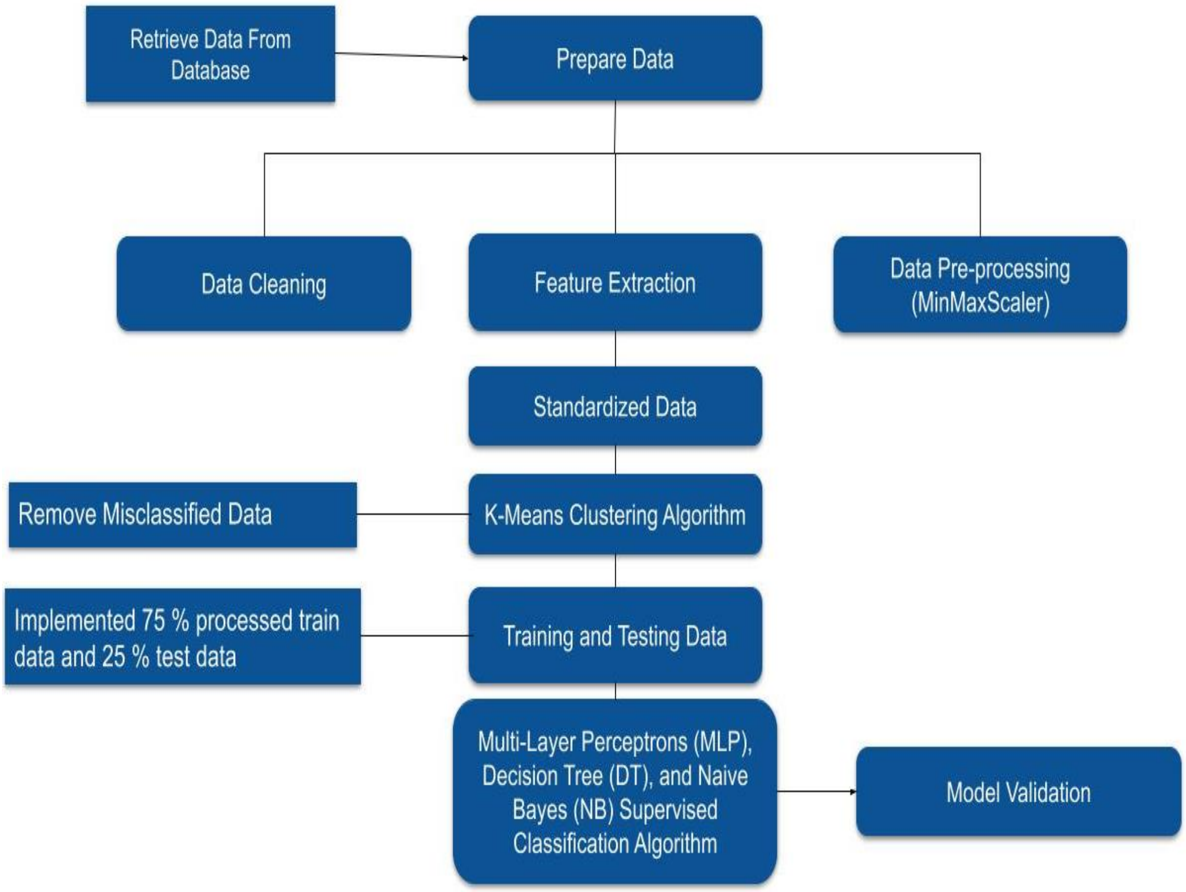

Figure1: Proposed Algorithm Steps in COVID-19 Prediction.

5. Data collection & analysis

5.1 Classification, Regression, And Clustering In Machine Learning

The systematic method of dividing systems into recognizable groupings and subcategories depending on their commonalities is called classification.. Many researchers used the concepts of classification, regression and clustering in Machine Learning to analyse & investigate the diseases [129], [129]–[133]. There are several sorts of categorizations in machine learning. MLP,DT, and NB Classifier. . Classifications comprise data that has been structured and annotated. For example, consider Figure 1, It displays a variety of categorization components used in a variety of processes.

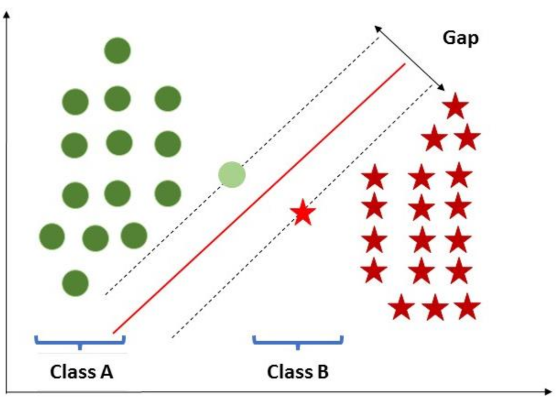

Figure2: Classification [134]

Because different regression models differ in terms of the type of interaction between dependent and independent variables, linear and nonlinear regression procedures are based on unsupervised & supervised learning techniques. It does regression tasks.

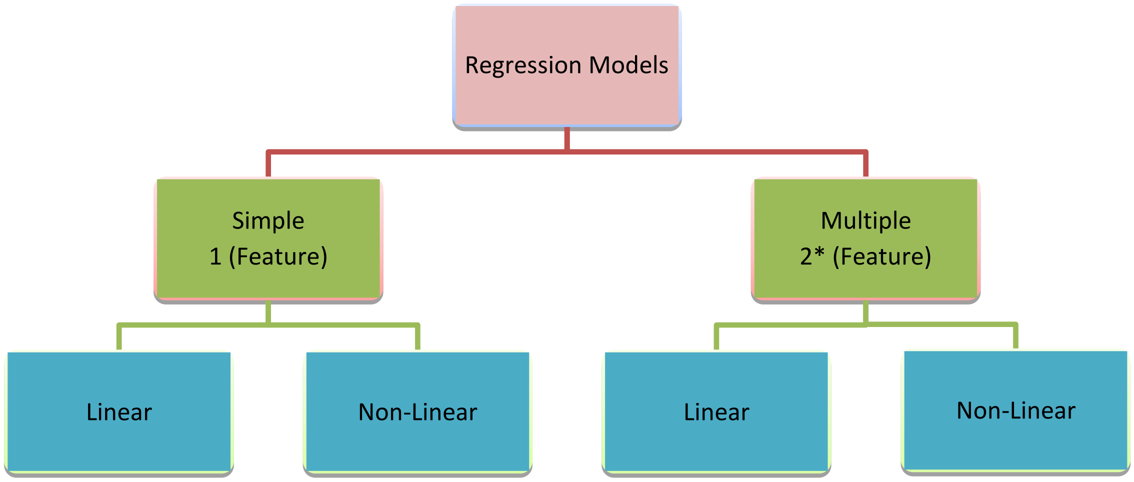

Figure3: Regression Models

As illustrated in figure3, Machine learning techniques employ both organized and unstructured data, as well as a variety of regression features. Both of these non-linear and linear regression incorporates the first and second properties of the regression model. Clustering is a particularly common kind of learning that is unsupervised, which has many uses across several sectors. A cluster is a group of related pieces of information that have undergone isolation and processing based on a data machine (ID). The various clusters are shown in Figure4.

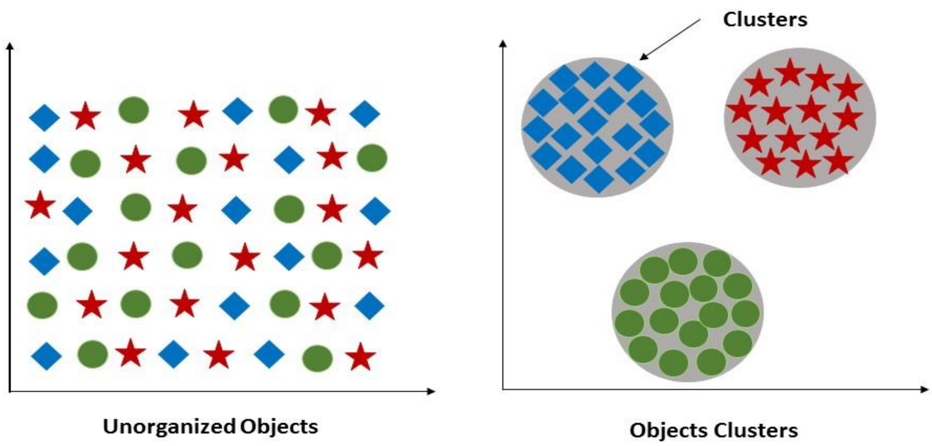

Figure4:Clustering [64]

5.2 Artificial Intelligence

AI is a branch of computer science that focuses on developing smart computers capable of performing tasks that typically require human intelligence. This field involves the creation of algorithms and models that enable computers to analyze data, make logical deductions, and generate predictions or conclusions. Artificial intelligence encompasses various domains such as robotics, ML, natural language processing, computer vision, and more. Its objective is to imitate and automate cognitive functions like decision-making, pattern recognition, and problem-solving.

5.3 Performance Analysis

High-level scripting languages like Python are frequently used in web development, databases, software, and machine learning methods. The GUI foundation of Anaconda Navigator -> Jupiter Notebook was used. We linked datasets together and ran the K-Means, DT, NB, and MLP algorithms using the Python computer language. The collection mostly consists of Covid-19 prediction diagnoses based on symptoms. This dataset contains tuples (rows) and 10-dimensional characteristics (columns) on the number 1048575. The identical dataset was used in each of the three independent programs, which we ran. The first program used K-Means and MLP procedures. The second program used Decision Tree method. The third program used the Naive Bayes process. All of these applications are active on the personal computer. The setup of the PC is mentioned below:

- Intel (R) Core (TM) i5-2520M, 2.00 GHz, second generation.
- The computer's operating system utilizes a 64-bit architecture along with a 4 GB RAM configuration.
- Window 10 Home.
- A 500 GB hard drive.

5.4 Data Collection

We have utilized a GB predictor trained using the LightGBM Python package. This package is a popular choice for implementing gradient boosting algorithms efficiently. To train our model, we divided our data into training and validation sets. The validation set was specifically used for early stopping, which is a technique employed during model training to prevent overfitting and find the optimal number of iterations or epochs. It helps to avoid training the model for too long, as that can lead to a decrease in generalization performance. In our case, we used the area under the receiver operating characteristic curve (auROC) as the performance measure. The auROC is a common evaluation metric for binary classification models, and it assesses the model's ability to distinguish between the positive and negative classes. Overall, we utilized a gradient-boosting predictor with the LightGBM package, employing early stopping with the auROC metric on the validation set to train and assess the performance of our model which is evident in many other studies [22], [65]–[67]. SHAP (Shapley Additive exPlanations) values are used to determine the main factors influencing the predictions made by complex models like artificial neural networks and gradient-boosting machines. SHAP values provide a way to measure the contribution of each feature or input variable to the model's output. They are particularly useful for understanding the importance of different features in complex models that are typically difficult to interpret. By calculating SHAP values, it becomes possible to identify the key features that drive the model's predictions. In this context, the SHAP values were specifically used to analyze artificial neural networks and gradient-boosting machines. These models are known for their complexity and black-box nature, making it challenging to understand the specific factors that influence their predictions. However, by employing SHAP values, researchers or analysts can gain insights into the relative importance of different features and better comprehend how these models make their predictions [135]–[138]. Random bootstrap sampling technique is commonly used method to obtain reliable estimates of relevant coefficients. This approach helps mitigate sampling bias and ensures that our estimates are not influenced by under- or over-weighting of specific groups of observations. By randomly selecting subsets of the original data and creating multiple bootstrap samples, we can generate a more robust understanding of the coefficients we are interested in estimating. This technique enhances the accuracy and reliability of our results by accounting for variations and uncertainties within the data [139]–[143]. Gathered data from the Kaggle website, which contains a large number of machine learning datasets, and the Israeli Ministry of HealthCovid-19 dataset, which was previously used by [144]. The dataset includes 508513 male and 537182 female sample patients from Israeli, Israeli residents who were observed for Covid-19 diagnosisand total 10 attributes (every attribute on behalf of medical diagnosis measures) and with one target class (which signifies the status of each tested separate) and in this dataset a total of523922tested negative cases and 524653tested positive cases [68], [69]. Few other authors also used the same data source to get the dataset for the Covid-19 disease diagnosis & analysis by using ML approach [145]–[148]. Table 1 briefly describes the characteristics and variables included in the implementation dataset for this study, which wasIsraeli Ministry of Health Covid-19 Dataset.

Table 1: Original Israeli Ministry of Health Covid-19 Dataset

Date time	**Cough**	**Fever**	**Sore Throat**	**ShortnessOf Breath**
11/12/2020	0	0	0	0
11/12/2020	0	1	0	0
11/12/2020	0	0	0	0
11/12/2020	0	0	0	0
11/12/2020	0	1	0	0
Headache	**Corona Result**	**Age60andAbov**	**Gender**	**Test Indication**
0	negative	no	male	other
0	negative	no	male	other
0	negative	yes	female	other
0	negative	no	male	other
0	negative	no	male	contact with confirmed

It also examines data normalization, preprocessing, simulation and induction needs, essential criteria, complication considerations, identified construct post-processing, simulations, and instructional models, the most important numerous steps by steps defined. Initially, data load from database subsequently prepare data and when data preparation process complete, then data preprocessing, standardized data. The cleaned and pre-processed data display table 2. The mixed data chart is shown in Figure 4withoutK-Means implementation. The X value colour circle is purple, and the second Y value colour circle is yellow.

Table 2: Pre-processed Israeli Ministry of Health Covid-19 Dataset

Cough	**Fever**	**Sore Throat**	**ShortnessOf Breath**
0.209985	0.050759	0.684658	0.145202
0.876941	0.069034	0.230935	0.363428
0.850973	0.061921	0.336166	0.353592
0.363601	0.646058	0.919711	0.548252
0.814712	0.867479	0.977511	0.889862
Headache	**Corona Result**	**Gender**	**Test Indication**
0.959292	0.858495	0.826742	0.463016
0.320959	0.597697	0.416783	0.002954
0.576014	0.553432	0.739722	0.838980
0.861257	0.628242	0.376887	0.510901
0.917373	0.261651	0.056288	0.412892
	Test Indication		
	0.852285		
	0.722472		
	0.163775		
	0.199961		
	0.035584		

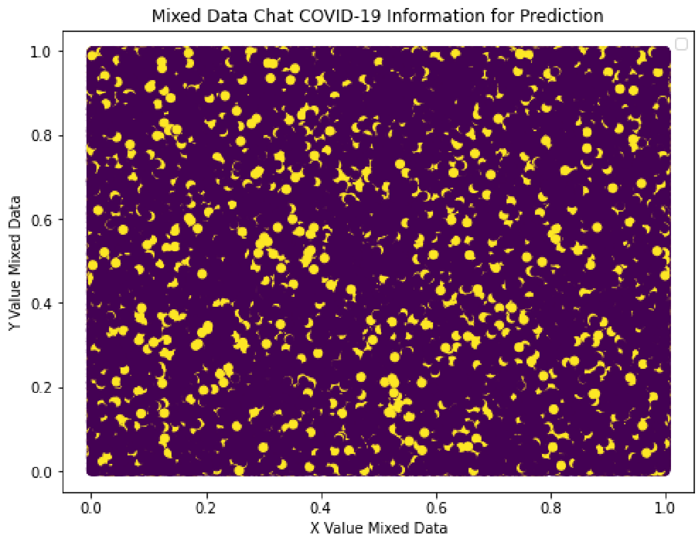

Figure 4: K-Means Earlier Execution Mixed Data Chart

5.5 K-Means Clustering Algorithm

The most popular kind of unsupervised learning, known as clustering, has a wide range of uses and widespread adoption in several fields. In order to create a set of data identified as clustering, information must be broken up and processed by a computer.Every cluster is assigned a distinctive identification number for identification purposes.The unsupervised K-means algorithm is a method of machine learning. This method categorises data as mixed or unstructured. The dataset starts with a first group of randomly chosen mean values, which acts as the beginning point for each following group, and then performs repetitive computations to optimise the position of the centre values [149]. The K-means algorithm operates on the following fundamental principles:

1. Determine the number of K clusters.
2. Compute the centroids by first sorting the dataset and then picking the K information for the centroids at random without substitution.
3. Recognizing fissures until there is minimal growth inside the centroids. In other words, the process of focusing information on clusters does not change.
4. Compute the quantity of the shaped distances between the details' focus facts and all the centroids.
5. Label each piece of information by pointing to the nearest whole number (centroidsignificance).
6. Determine the centroids for the clusters by calculating the cruel of all information focuses assigned to each cluster.
7. End.

These are a few mathematical separation strategies or functions that may be used to any algorithm. This method, like the Euclidean, Manhattan, and Hamming distances, is based on classification. Among these approaches are the following:

Euclidean $$\sqrt{\sum_{i=1}^{k}(xi-yi)2}$$

Manhattan $$\sum_{i=1}^{k}|xi-yi|$$

Minkowski $$\left(\sum_{i=1}^{k}(|xi-yi|)q\right)1/q$$

In this preprocessing, we included a standardized dataset and utilized it to generate the mixed data presentations. K-Means is a technique for data cleaning, and data preprocessingof such datasets, increasing interpretability while minimizing information loss, using K-Means two clusters.Within the K-Means clustering technique, each data point is allocated a probability or possibility score, indicating the degree of likelihood that it belongs to a particular cluster. It accomplishes this by successively producing new, highly asymmetrical variables. This reasoning was instantly used on the data matrix to produce the membership matrix that shows how closely connected each cluster is to each sample. This method essentially uses a clustering strategy, as shown by the ten-dimensional dataset, K-Means algorithm, and centroid clustering values. The Ministry of Health Covid-19 dataset, which has nine dimensions, is combined with the Clusters dataset. The Clusters dataset includes three features derived from the preprocessing values of the Ministry of Health Covid-19 dataset, as well as one feature representing the target property, which is the cluster number (see Table 3).Below is a concise definition of the K-Means Clustering Centroid Value. The structured data obtained from converting the unstructured dataset is represented by two clusters, visualized in Figure 5. The centroid values for the clusters, represented by the colors green and red, are indicated by the yellow asterisk (*) symbol. This information can be found in Table 6 illustrates a line chart representing the sum of squared error determined by the K-Means algorithm.

K-Means Clustering Centroid Value.

array([[0.50179588, 0.50052385, 0.50155221, 0.49727497, 0.49845892,0.50047932, 0.50111732, 0.25017634, 0.49937262],
[0.49820804, 0.50010101, 0.49865365, 0.50206582, 0.50129637, 0.50033012, 0.49813375, 0.75055154, 0.50049609]])

Table 3:K-Means Two ClustersPreprocessedPre-Processed Israeli Ministry of Health Covid-19 Dataset

Cough	**Fever**	**Sore Throat**	**ShortnessOf Breath**	**Test Indication**
0.209985	0.050759	0.684658	0.145202	0.463016
0.876941	0.069034	0.230935	0.363428	0.002954
0.850973	0.061921	0.336166	0.353592	0.838980
0.363601	0.646058	0.919711	0.548252	0.510901
0.814712	0.867479	0.977511	0.889862	0.412892
Headache	**Corona Result**	**Gender**	**Test Indication**	**Clusters**
0.959292	0.858495	0.826742	0.852285	0
0.320959	0.597697	0.416783	0.722472	0
0.576014	0.553432	0.739722	0.163775	1
0.861257	0.628242	0.376887	0.199961	1
0.917373	0.261651	0.056288	0.035584	0

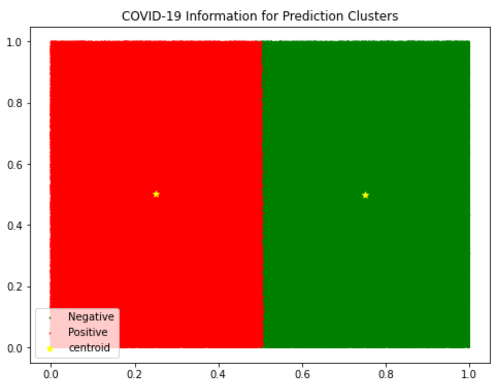

Figure 5: K-Means Two Clusters

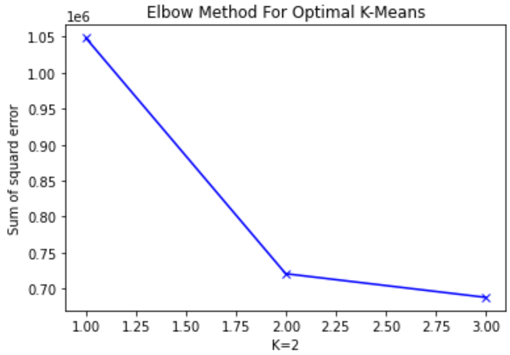

Figure 6: K-Means Sum of Squared Error Line Chart

Utilizing accuracy, f-measure, and recall we relate the development of CFD exhausting-equally clustering algorithms. The classification of COVID-19 results as legitimate, suspicious, or predicted is determined by evaluating a suspicion score, which measures the degree of deviation from normal patterns. This scoring mechanism assesses the extent of abnormality to categorize the COVID-19 outcome.

5.6 Multi-Layer Perceptrons (MLP) Algorithm

An advanced optimization algorithm called the Multilayer Perceptron (MLP) is composed of multiple perceptrons. It consists of an input layer that receives input data, an output layer that generates judgments or estimates based on the input, and an arbitrary number of hidden layers that serve as the MLP's computational power. By varying the number of hidden layers, the MLP is capable of approximating any continuous function [150], [151]. In cases where datasets are not conditionally independent, the MLP overcomes this challenge by employing participants to develop machine learning and prediction models with a more flexible and complex framework. This approach, often used in supervised learning, addresses challenges related to difficult data patterns and enables scientific advancements in various fields. Some of these approaches are mentioned below:

Sigmoid	$\mathbf{S}\,(z) = \frac{\mathbf{1}}{(\mathbf{1} + \mathbf{e}^{-\mathbf{z}})}$
Linear Regression	$y = e\hat{}(b0 + b1 * x) / (1 + e\hat{}(b0 + b1 * x))$
Cost Linear Regression	$(Cost\ (h\theta(x), y)) = -log\ (h\theta(x)), if\ y = 1\ and$ $(Cost\ (h\theta(x), y)) = -\ log\ (1 - h\theta(x)), if\ y = 0$
Nonlinear Regression	$\mathbf{Y} = \mathbf{f}(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}$

The MLPalgorithm main steps exertion as follows:

1. Similar to a perceptron, the Multilayer Perceptron (MLP) utilizes partial derivatives of input data and parameters between input and hidden layers to compute values in the hidden layer. However, unlike an activation function, these values are not increased.
2. MLPs employ activation functions like rectified linear units (ReLU), sigmoid, or tanh in each of their hidden layers. These functions transform the calculated outputs to the visible layer.
3. Once the desired output is generated in the hidden layer through the activation function, partial derivatives are computed with respect to the relevant values and propagated to the next layer within the MLP.
4. This process is repeated for additional layers until the desired outcome is achieved.
5. The calculated results are then utilized for further analysis or decision-making.

The MLP prediction labels the old data values and predicts the value of the new data, where we try to make your predictions fit the labels during preparation,figure 7 shows the result of that MLPconfusion matrix.

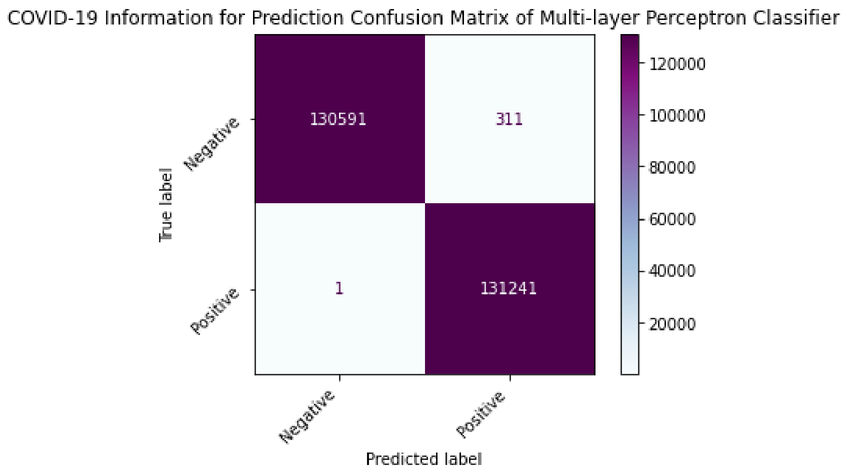

Figure 7: Confusion Matrix Multi-Layer Perceptrons (MLPs) Algorithm

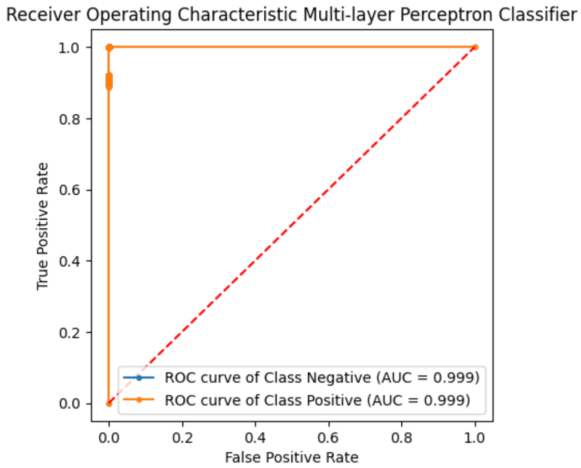

Figure 8:Multi-Layer Perceptrons (MLP) Receiver Operating Characteristic (ROC Curve)

In the context of this paper, the confusion matrix is represented as [[A, B], [C, D]]

- A represents the count of accurate predictions classifying an example as negative
- B signifies the count of inaccurate predictions labeling an example as positive
- C denotes the count of negative extrapolations that are incorrect
- D indicates the count of correct predictions identifying an example as positive.

By showing the Receiver Operating Characteristic curve for a no-skill classifier on a fictitious dataset, we can demonstrate this. The resulting MLP algorithm ROC Curve picture 8 illustrates outcome.

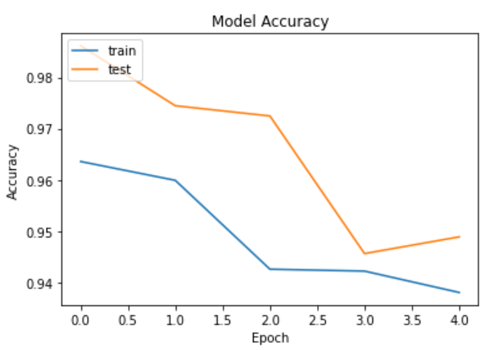

Figure 9: Model Accuracy Multi-Layer

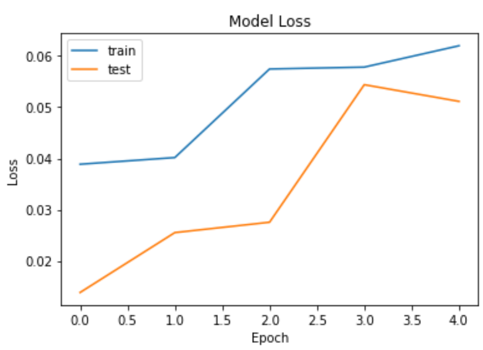

Figure 10: Model Loss Multi-Layer Perceptrons

Perceptrons (MLPs) Algorithm (MLPs) Algorithm

In this paper, we used the concept of the ROC curve to analyze our model. This will gain the accuracy of the estimated prediction method while frequently meeting the patterns of covid-19 prediction. In this study, we examined our model using the model accuary and model loss concepts. This will improve the accuracy of the approximated prediction approach while also fulfilling the patterns of covid-19 prediction on a regular basis.Figure 9, 10 displays the resultmodel accuracy and model loss multi-Layer perceptrons (MLPs) algorithm.

5.7 Decision Tree Algorithm

The Decision Tree method falls under the umbrella of supervised ML methods. In addition, the decision tree approach may be utilized to beat classification techniques applications. The purpose of creating a Tree Structure is to design a development model that predicts the category or values of the target attribute by learning fundamental choice rules from past results (training data). In Decision Trees, we begin at the root of the tree to generate a data point for a data. The importance of the record's quality is balanced against the quantities of the fundamental property. We proceed towards the next nodes involves the comparison by following the branches that equivalent to that outcome. These are some scientific methods or functions that may be utilized in conjunction with any algorithm [152]. This approach, like the Information Gain and Entropy methods, is established on the creation of a DT and classification. The following are some of these approaches:

Information Gain	$\textbf{Information Gain} = \mathbf{E(Y) - E(Y|X)}$
Entropy	$\boldsymbol{Entropy(s) = -P(+)log2\,P(+) - P(-)log2\,P(-)}$

The DTalgorithmsignificant steps work as follows:

1. It started with S, the initial root node.
2. The procedure repeats over the comparatively duplicate variable of the sequence, evaluating the Entropy (H) and Information Gain (IG) of this capability with each iterative process.
3. The feature with the shortest entropy or greatest information gain is instead selected.
4. The measurement S is instead partitioned by such desired characteristic to provide a number of observations.
5. The approach is repeated for each segment, reflecting only qualities which have never been accepted previously.
6. End.

With the use of the decision tree algorithm,Themulti-layer perceptrons prediction labels the old data values. It predicts the value of data, where we try to make your predictions fit the labels during preparation with the use of algorithm decision tree algorithm figure.11 shows the result of that confusion matrix.

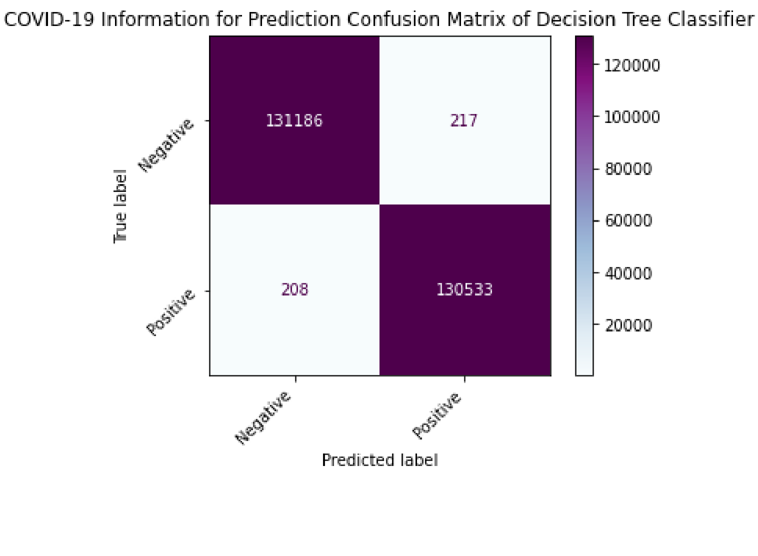

Figure 11: Confusion Matrix Decision Tree Algorithm

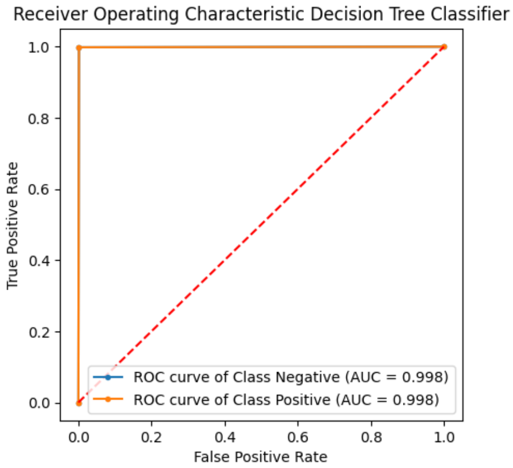

Figure 12: Decision Tree Receiver Operating Characteristic (ROC Curve)

Assuming that ourdecision tree model sounds fitted in this, the confusion matrix helped us calculate the predicted labels of our detection prediction mentioned in figure 11 & 12.

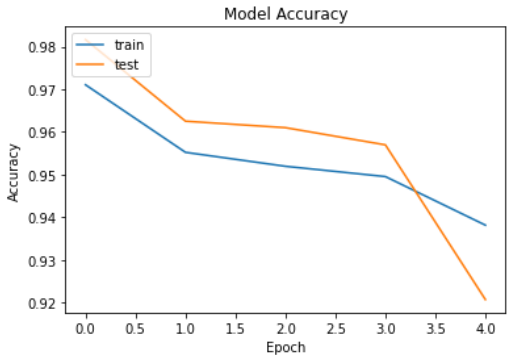

Figure 13: Model Accuracy Decision TreeAlgorithm

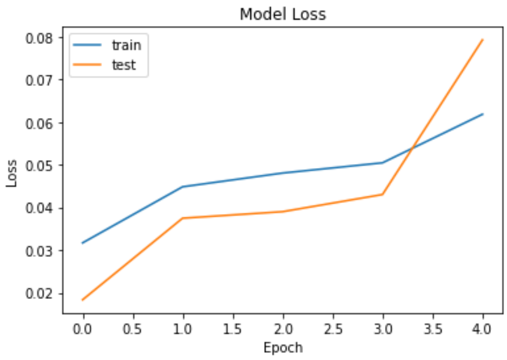

Figure 14: Model Loss Decision TreeAlgorithm

In this work, we evaluated our model using the model accuary and model loss concepts. This will improve the approximated prediction strategy's accuracy while also fulfilling the covid-19 forecast patterns on a regular basis.Figure13, 14presentations the resultmodel accuracy and model loss decision tree algorithm.

5.8 Naive Bayes Algorithm

A supervised ML method built on the Bayes theorem, Naive Bayes (NB) algorithm manages categorization issues. This technique is typically employed in sentiment analysis with a sizable train data set, and it is a probabilistic and practical classification strategy that helps in the construction of quick ML algorithms' capacity to provide precise forecasts. These are methods or statements from science that may be applied to enhance any algorithm [152]. This method is classification-based, like the Bayes Theorem method. These viewpoints are as follows::

Bayes Theorem Formula

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)} \quad (10)$$

The probability of an incidence is calculated by the Naive Bayes classifier in the following steps:

1. Compute the statistical significance for every one of the target class indicated.
2. Estimate the significant chance to every parameter per each category.
3. Calculate the conditional probabilities using the Bayes Formula and these values.
4. Using the Probabilistic Technique and enter inputs data, compute the conditional probability.
5. Evaluate if the category has an increased probability, provided that such input relates to the data with a high possibility.
6. End.

With the NB algorithm, the multi-layer perceptronspredictionlabels the previous values. The NB method is employed to predict the values of the data, and during the training phase, efforts are made to align the forecasts with the corresponding labels. The resulting confusion matrix is presented in Figure 15.

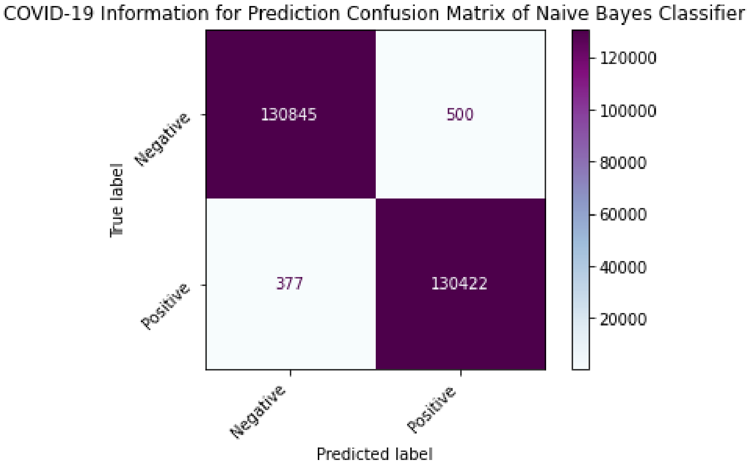

Figure 15: Confusion Matrix Naive BayesAlgorithm

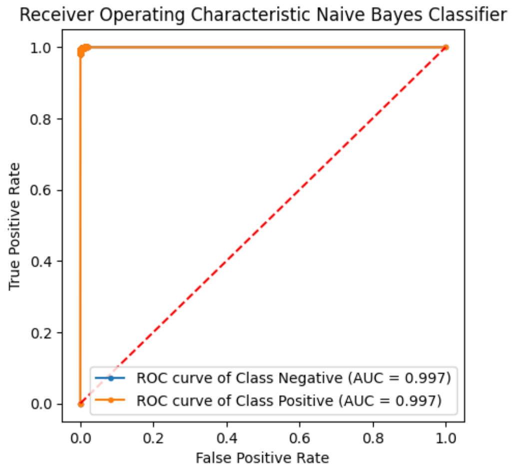

Figure 16: Naive BayesReceiver Operating Characteristic (ROC Curve)

The best thing about this confusion matrix is that is doesn’t mislead the outcome. It illustrates a classifier model's systematic performance when its filtering inception is altered. In diagnostic judgement, ROC analysis is inextricably linked to cost/benefit evaluation, as seen in NB ROC Curvefigure 16.

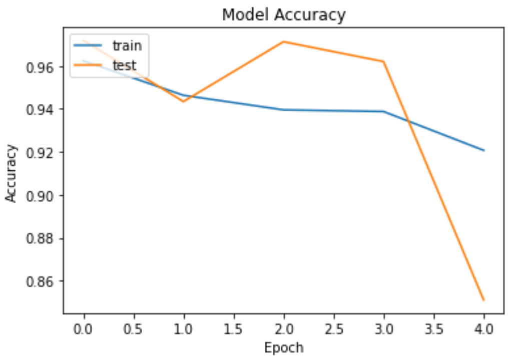

Figure 17: Model AccuracyNaive BayesAlgorithm

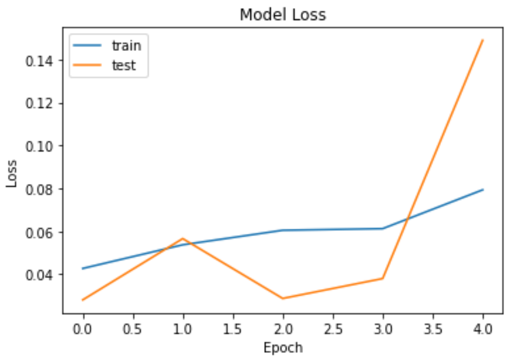

Figure 18: Model Loss Naive BayesAlgorithm

In this study, we assessed the performance of our model by considering the framework's accuracy and modeling loss metrics. By doing so, we aimed to enhance the accuracy of the approximate prediction strategy while ensuring its consistency with the patterns observed in covid-19 predictions. The outcomes of the naive Bayes algorithm for model accuracy and model loss can be observed in Figures 17 and 18, respectively.

6. Results and discussion

Machine learning is an empirical technique that teaches algorithm to solve challenges instead of being supervised ML. DL is currently the ML champion, thanks to better techniques, computer power, and large data sets. Nevertheless, conventional ML technique continue to play an important role in the industry. The paper introduces a novel approach that combines the integration of K-Means with other Supervised Machine Learning Algorithms (SMLA) to create a powerful classifier. The researchers conducted a comparative study to evaluate the performance of their approach against other commonly used SMLA. The integration of K-Means is a key component of their approach, which allows for the identification of clusters within the data. This clustering helps in grouping similar instances together, enabling more effective classification. By combining K-Means with SMLA, the researchers aimed to enhance the classifier's accuracy and performance. To determine the best classifier, the researchers compared the performance of their integrated approach with other established SMLA. They likely used various metrics such as accuracy, precision, recall, or F1 score to assess the performance of each classifier. By conducting this comparative study, they aimed to identify which classifier yielded the highest performance on their dataset. Overall, the paper proposes a new approach that integrates K-Means with SMLA for classification tasks. The researchers performed a comparative study to determine the best classifier, and their findings provide insights into the effectiveness of their approach compared to other methods. The SMLA techniques used in this studyare Multi-Layer Perceptrons (MLPs), Decision Trees (DT), and Naive Bayes. In the next step, we combineed the algorithms to check the most accurate combination over covid-19 diagnosis prediction based on symptoms.The highest accuracy we gained using our algorithms' combination was through K-Means, Multi-Layer Perceptrons (MLPs). Next we acquire the second-ranked combination which was K-Means, Decision Tree. Finally, K-Means, Naïve Bayes was ranked third.

Table 4: Combination of Algorithms Model Accuracy Covid-19 Diagnosis Prediction Based On Symptoms

Combination of Algorithms	**Accuracy**
K-Means,Multi-Layer Perceptrons (MLPs)	99.8809814453125 %
K-Means, Decision Tree (DT)	99.83787536621094%
K-Means,Naïve Bayes (NB)	99.66545104980469 %

Table 5: Combination of Algorithms Parameter Score Covid-19 Diagnosis Prediction Based On Symptoms

S/#.	**Parameter Score (%)**	**K-Means, Multi-Layer Perceptrons (MLPs)**	**K-Means, Decision Tree (DT)**	**K-Means, Naïve Bayes (NB)**
1	Accuracy	0.99880981	0.99837875	0.99665451

2	Precision	0.99880981	0.99837875	0.99665451
3	Recall	0.99880981	0.99837875	0.99665451
5	Sensitivity	0.99762417	0.99834859	0.99619323
6	Specificity	0.99999238	0.99840906	0.99711771
7	F1-Score	0.99880981	0.99837875	0.99665451

The combination of our implemented algorithm, which is K-Means, Multi-Layer Perceptrons (MLPs), has reached at its maximum outcome in terms of accuracy. However, the combination between K-Mean and DT was ranked second, followed by our third algorithm consisting of K-Means and Naïve Bayes Algorithm and table 4, and 5 briefly defined results.

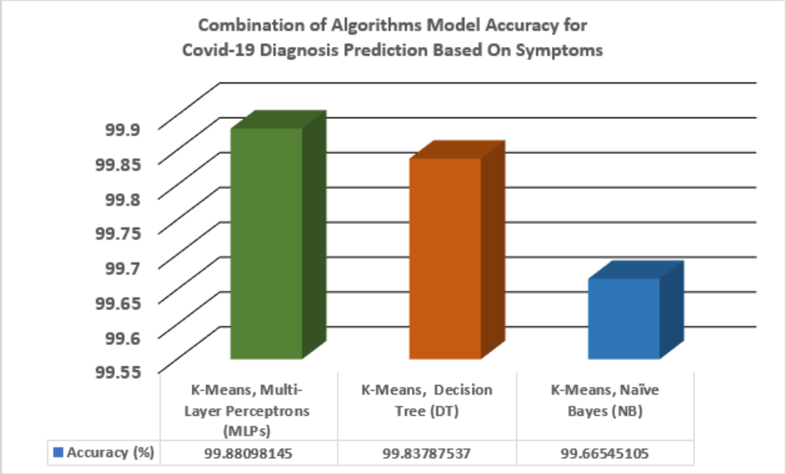

Figure 19: Combination of Algorithms Model Accuracy Covid-19 Diagnosis Prediction Based On Symptoms

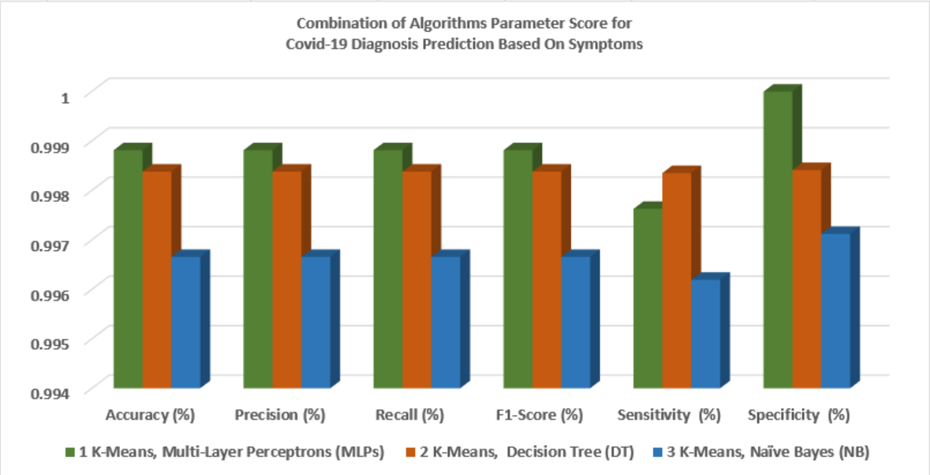

Figure20: Combination of AlgorithmsParameter ScoreCovid-19 Diagnosis Prediction Based On Symptoms

As per the results, the accuracy graph also shows the combinations' prediction at its maximum mentioned in Figure 19. Depending on the context, there may be adjustments or reductions made to the accuracy metric. Figure 20 illustrates that accuracy, precision, recall, sensitivity, f1-Score , and specificityare optimized and exhibit their highest levels of correctness

7. Conclusion

The novelty of this research lies in the comprehensive exploration and utilization of multiple algorithms, namely K-Means Clustering, Multi-Layer Perceptrons (MLPs), Decision Tree, and Naive Bayes, for COVID-19 case detection. By leveraging these algorithms and the Israeli Ministry of Health COVID-19 dataset, the study achieved highly accurate results. The significant contributions of K-Means and Multi-Layer Perceptrons (MLPs) to the overall accuracy highlight their potential effectiveness in COVID-19 case identification. Furthermore, the application of the K-Means methodology to enhance accuracy adds novelty to the research approach. We utilized the publicly released COVID-19 dataset from the Israeli Ministry of Health to classify and cluster the data. The K-Means methodology was employed to enhance accuracy. Combining the K-Means and Multi-Layer Perceptrons (MLPs) yielded an accuracy of 99.8809814453125%. Similarly, combining the K-Means and Decision Tree resulted in an accuracy of 99.83787536621094%, and combination of K-Means with the Naive Bayes Algorithm produced an accuracy of 99.66545104980469%. The future direction of this work involves further exploration and refinement of the employed algorithms for COVID-19 case detection. Building upon the success of combining K-Means and Multi-Layer Perceptrons (MLPs), future research could investigate the integration of other advanced machine learning techniques to enhance accuracy. Additionally, the study could expand its dataset sources beyond the Israeli Ministry of Health, considering diverse and comprehensive data to validate the algorithmic approaches in different contexts and regions.

8. Future Implications

This study built and examined various ML & DL algorithms of supervised & unsupervised learning by creating a classifier depending on a predefined model to estimate COVID-19. K-Means, Decision Tree, Naive Bayes and Multi-Layer Perceptron algorithms are used for the data analysis. The same data sets can be used to explore other ML & DL algorithms in the prediction of COVID-19 and other diseases. The other data sets can be used for the same analysis and/or other analysis in COVID-19 and/or other diseases.

9. Limitations

Since the specific data is used for predicting COVID-19 diagnosis by ML and DL approaches, therefore the results are limited to the scope of used data. The specific ML & DL algorithms are used for the analysis therefore the results are limited to the scope of these ML & DL algorithms

10. Acknowledgement

The main author appreciate the collaborations of all the co-authors appreciate their continuous support. We would like to thank to the data providing agencies and field specialist for their devoted cooperation throughout this project.

11. Conflict of Interest

The current study's authors have no conflicting interests.

REFERENCES

[1] Centers for Disease Control and Prevention. "CDC. COVID-19 and your health." (2020).

[2] Amara, Sara Aicha, Estefany Daniella Díaz, Lakshmi Krishna Menon, Priyanka Singh, Liudmila Rozanova, and Antoine Flahault. "COVID-19 Outbreak Management and Vaccination Strategy in The United States of America." *Epidemiologia* 2, no. 3 (2021): 426-453.

[3] Simoes, Eduardo J., and Jeannette Jackson-Thompson. "The United States public health services failure to control the coronavirus epidemic." *Preventive Medicine Reports* (2022): 102090.

[4] Benchimol-Elkaim, Brandon, Scott Dryden-Peterson, Donald R. Miller, Howard K. Koh, and Alan C. Geller. "Oral Antiviral Therapy Utilization Among Adults with Recent COVID-19 in the United States." *Journal of General Internal Medicine* (2023): 1-5.

[5] Benchimol-Elkaim, Brandon, Scott Dryden-Peterson, Donald R. Miller, Howard K. Koh, and Alan C. Geller. "Oral Antiviral Therapy Utilization Among Adults with Recent COVID-19 in the United States." *Journal of General Internal Medicine* (2023): 1-5.

[6] Nurmaini, Siti, Alexander Edo Tondas, Radiyati Umi Partan, Muhammad Naufal Rachmatullah, Annisa Darmawahyuni, Firdaus Firdaus, Bambang Tutuko, Rachmat Hidayat, and Ade Iriani Sapitri. "Automated Detection of COVID-19 Infected Lesion on Computed Tomography Images Using Faster-RCNNs." *Engineering Letters* 28, no. 4 (2020).

[7] Tostmann, Alma, John Bradley, Teun Bousema, Wing-Kee Yiek, Minke Holwerda, Chantal Bleeker-Rovers, Jaap Ten Oever et al. "Strong associations and moderate predictive value of early symptoms for SARS-CoV-2 test positivity among healthcare workers, the Netherlands, March 2020." *Eurosurveillance* 25, no. 16 (2020): 2000508.

[8] Punn, Narinder Singh, Sanjay Kumar Sonbhadra, Sonali Agarwal, and Gaurav Rai. "Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques." *arXiv preprint arXiv:2005.01385* (2020).

[9] Mei, Xueyan, Hao-Chih Lee, Kai-yue Diao, Mingqian Huang, Bin Lin, Chenyu Liu, Zongyu Xie et al. "Artificial intelligence–enabled rapid diagnosis of patients with COVID-19." *Nature medicine* 26, no. 8 (2020): 1224-1228.

[10] Feng, Cong, Lili Wang, Xin Chen, Yongzhi Zhai, Feng Zhu, Hua Chen, Yingchan Wang et al. "A Novel triage tool of artificial intelligence-assisted diagnosis aid system for suspected COVID-19 pneumonia in fever clinics." *MedRxiv* (2020): 2020-03.

[11] Hassan, Faisal, Naseem Afzal Qureshi, Muhammad Zohaib Khan, Muhammad Ali Khan, Abdul Salam Soomro, Aisha Imroz, and Hussain Bux Marri. "Performance evolution for sentiment classification using machine learning algorithm." *Journal of Applied Research in Technology & Engineering* 4, no. 2 (2023): 97-110.

[12] Amor, Nahla Ben, Salem Benferhat, and Zied Elouedi. "Naive bayes vs decision trees in intrusion detection systems." In *Proceedings of the 2004 ACM symposium on Applied computing*, pp. 420-424. 2004.

[13] Astuti, Yuli, Sri Ngudi Wahyuni, Dina Maulina, and Fajar Muhammad Sidiq. "The Data Leakage Sentiment Analysis Using Naive Bayes Algorithm Based on Machine Learning Approach." In *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pp. 215-220. IEEE, 2022.

[14] Nayak, Madhu M., and Sumithra Devi Kengeri Anjanappa. "An Efficient Hybrid Classifier for MRI Brain Images Classification Using Machine Learning Based Naive Bayes Algorithm." *SN Computer Science* 4, no. 3 (2023): 223.

[15] Khamdamovich, Khamdamov Rustam, and Haydarov Elshod. "Detecting spam messages using the naive Bayes algorithm of basic machine learning." In *2021 International Conference on Information Science and Communications Technologies (ICISCT)*, pp. 1-3. IEEE, 2021.

[16] Tzanos, Georgios, Christoforos Kachris, and Dimitrios Soudris. "Hardware acceleration on gaussian naive bayes machine learning algorithm." In *2019 8th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, pp. 1-5. IEEE, 2019.

[17] Jordan, Vanessa. "Coronavirus (COVID-19): using signs and symptoms to diagnose COVID-19 in primary care." *Journal of Primary Health Care* 14, no. 2 (2022): 187-188.

[18] Wang, Sonia, Joel M. Gelfand, and Cassandra Calabrese. "Outpatient Management of COVID-19: A Primer for the Dermatologist." *Current Dermatology Reports* 11, no. 4 (2022): 318-327.

[19] Mølhave, Martin, Jane Agergaard, and Christian Wejse. "Clinical management of COVID-19 patients–an update." In *Seminars in Nuclear Medicine*, vol. 52, no. 1, pp. 4-10. WB Saunders, 2022.

[20] Shi, Honghao, Jingyuan Wang, Jiawei Cheng, Xiaopeng Qi, Hanran Ji, Claudio J. Struchiner, Daniel AM Villela, Eduard V. Karamov, and Ali S. Turgiev. "Big data technology in infectious diseases modeling, simulation, and prediction after the COVID-19 outbreak." *Intelligent Medicine* (2023).

[21] E. Taghizadeh and A. Mohammad-Djafari, “SEIR Modeling, Simulation, Parameter Estimation, and Their Application for COVID-19 Epidemic Prediction,” in Physical Sciences Forum, Multidisciplinary Digital Publishing Institute, 2022, p. 18.

[22] Zoabi, Yazeed, Shira Deri-Rozov, and Noam Shomron. "Machine learning-based prediction of COVID-19 diagnosis based on symptoms." *npj digital medicine* 4, no. 1 (2021): 3.

[23] Mohamadou, Youssoufa, Aminou Halidou, and Pascalin Tiam Kapen. "A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19." *Applied Intelligence* 50, no. 11 (2020): 3913-3925.

[24] Li, Lixiang, Zihang Yang, Zhongkai Dang, Cui Meng, Jingze Huang, Haotian Meng, Deyu Wang et al. "Propagation analysis and prediction of the COVID-19." *Infectious Disease Modelling* 5 (2020): 282-292.

[25] Liu, Ying, Albert A. Gayle, Annelies Wilder-Smith, and Joacim Rocklöv. "The reproductive number of COVID-19 is higher compared to SARS coronavirus." *Journal of travel medicine* (2020).

[26] Jin, Jian-Min, Peng Bai, Wei He, Fei Wu, Xiao-Fang Liu, De-Min Han, Shi Liu, and Jin-Kui Yang. "Gender differences in patients with COVID-19: focus on severity and mortality." *Frontiers in public health* (2020): 152.

[27] Sidor, Aleksandra, and Piotr Rzymski. "Dietary choices and habits during COVID-19 lockdown: experience from Poland." *Nutrients* 12, no. 6 (2020): 1657.

[28] O'Driscoll, Megan, Gabriel Ribeiro Dos Santos, Lin Wang, Derek AT Cummings, Andrew S. Azman, Juliette Paireau, Arnaud Fontanet, Simon Cauchemez, and Henrik Salje. "Age-specific mortality and immunity patterns of SARS-CoV-2." *Nature* 590, no. 7844 (2021): 140-145.

[29] Peckham, Hannah, Nina M. de Gruijter, Charles Raine, Anna Radziszewska, Coziana Ciurtin, Lucy R. Wedderburn, Elizabeth C. Rosser, Kate Webb, and Claire T. Deakin. "Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission." *Nature communications* 11, no. 1 (2020): 6317.

[30] Yuki, Koichi, Miho Fujiogi, and Sophia Koutsogiannaki. "COVID-19 pathophysiology: A review." *Clinical immunology* 215 (2020): 108427.

[31] Cascella, Marco, Michael Rajnik, Abdul Aleem, Scott C. Dulebohn, and Raffaela Di Napoli. "Features, evaluation, and treatment of coronavirus (COVID-19)." (2020).

[32] Singh, Satwinder, Harpreet Kaur, Rubal Kanozia, and Gurpreet Kaur. "Empirical Analysis of Supervised and Unsupervised Machine Learning Algorithms with Aspect-Based Sentiment Analysis." *Applied Computer Systems* 28, no. 1 (2023): 125-136.

[33] Kalwar, Muhammad Ahmed, Shakil Ahmed Shaikh, Muhammad Ali Khan, Muzamil Hussain Wadho, Sarmad Ali Khaskheli, Liaquat Ali Rahoo, and Hussain Bux. "HOSTILE BEHAVIOUR OF DRIVERS ON THE ROADS DURING ROUTINE TRAFFIC SITIATIONS IN AN EDUCATIONAL CITY."

[34] Rajput, Shahzain, Muhammad Ali Khan, Sajid Samejo, Ghulam Murtaza, and Rao Atif Ali. "Productivity Improvement by the Implementation of lean manufacturing practice (takt time) in an automobile assembling plant." In *Proceedings of the International Conference on Industrial Engineering and Operations Management Dubai, UAE*, pp. 1618-1619. 2020.

[35] Iftikhar, Z., M. A. Khan, R. Kumar, K. Bux, and A. Haseeb. "Productivity Improvement of Garments Industry by Assembly Line Technique of Lean Manufacturing." In Proceedings (Abstract) of the International Conference on Industrial & Mechanical Engineering and Operations Management Dhaka, Bangladesh, December 26-27, vol. 908. 2021.

[36] Haseeb, Abdul, Muhammad Ali Khan, Shakeel Ahmed Shaikh, Zohaib Iftikhar, Ramesh Kumar, Karim Bux, and Arshia Naz. "Performance optimization of pillow sheet set sewing assembly line by SAM analysis and lean manufacturing techniques of method study & work measurement." *Journal of Applied Research in Technology & Engineering* 4, no. 1 (2023): 1-12.

[37] Jaleel, N., M. A. Khan, M. Jamal, M. Safeeruddin, M. M. Shajee, and U. Mughal. "Productivity Improvement by Lean Methodologies at Dyeing & Printing Plant." In *Proceedings (Abstract) of the International Conference on Industrial & Mechanical Engineering and Operations Management Dhaka, Bangladesh, December 26-27*, vol. 905. 2021.

[38] Iftikhar, Zohaib, Ramesh Kumar, Karim Bux, Abdul Haseeb, M. A. Khan, A. Naz, and A. S. Soomro. "Lean Manufacturing Tools and Techniques for the Productivity Improvement in Assembly Lines Operations of Industries." *International Research Journal of Modernization in Engineering Technology and Science* 4, no. 7 (2022): 4554-4562.

[39] Zaman, Fahim Uz, Mansoor Ahemed Khuhro, Kamlesh Kumar, Naadiya Mirbahar, Muhammad Zohaib Khan, and Ayesha Kalhoro. "Comparative Case Study Difference Between Azure Cloud SQL and Atlas MongoDB NoSQL Database." *International Journal* 9, no. 7 (2021).

[40] Khan, Muhammad Zohaib, Fahim Uz Zaman, Muhammad Adnan, Aisha Imroz, Mahira Abdul Rauf, and Zuhaib Phul. "Comparative Case Study: An Evaluation of Performance Computation Between SQL And NoSQL Database." *Journal of Software Engineering* 1, no. 2 (2023): 14-23.

[41] M. Z. Khan et al., "Comparative case Study : An Evaluation of Performance Computation between Support Vector Machine , K-Nearest Comparative Study : Evaluation of Performance Computation Between Support Vector Component Analysis," J. Tianjin Univ. Sci. Technol., no. April, 2022, doi: 10.17605/OSF.IO/HK3SF.

[42] M. Z. Khan et al., "Comparative case Study : An Evaluation of Performance Computation between Support Vector Machine , K-Nearest Comparative Study : Evaluation of Performance Computation Between Support Vector Component Analysis," J. Tianjin Univ. Sci. Technol., no. April, 2022, doi: 10.17605/OSF.IO/HK3SF.

[43] Arain, Muhammad Saleem, Muhammad Ali Khan, and Muhammad Ahmed Kalwar. "Optimization of Target Calculation Method for Leather Skiving and Stamping: Case of Leather Footwear Industry." *International Journal of Education and Management Studies* 7, no. 1 (2020): 15-30.

[44] Kalwar, Muhammad Ahmed, and Muhammad Ali Khan. "Increasing performance of footwear stitching line by installation of auto-trim stitching machines." *Journal of Applied Research in Technology & Engineering* 1, no. 1 (2020): 31-36.

[45] Kalwar, Muhammad Ahmed, and Muhammad Ali Khan. "Optimization of Procurement & Purchase Order Process in Foot Wear Industry by Using VBA in Ms Excel." *International Journal of Business Education and Management Studies* 5, no. 2 (2020): 80-100.

[46] Kalwar, M. A., S. A. Shaikh, M. A. Khan, and T. S. Malik. "Optimization of Vendor Rate Analysis Report Preparation Method by Using Visual Basic for Applications in Excel (Case Study of Footwear Company of Lahore)." In *Proceedings of the International Conference on Industrial Engineering and Operations Management*, pp. 26-27. 2020.

[47] Kalwar, Muhammad Ahmed, and Muhammad Ali Khan. "Optimization of Procurement & Purchase Order Process in Foot Wear Industry by Using VBA in Ms Excel." *International Journal of Business Education and Management Studies* 5, no. 2 (2020): 80-100

[48] Kalwar, Muhammad Ahmed, Hussain Bux Marri, and Muhammad Ali Khan. "Performance Improvement of Sale Order Detail Preparation by Using Visual Basic for Applications: A Case Study of Footwear Industry." *International Journal of Business Education and Management Studies* 3, no. 1 (2021): 1-22.

[49] Khan, Muhammad Ali, Muhammad Ahmed Kalwar, Abdul Jabbar Malik, Tanveer Sarwar Malik, and Abdul Karim Chaudhry. "Automation of Supplier Price Evaluation Report in MS Excel by Using Visual Basic for Applications: A Case of Footwear Industry." *International Journal of Science and Engineering Investigations* 10, no. 113 (2021): 49-60.

[50] Khan, Muhammad Ali, Muhammad Ahmed Kalwar, and Abdul Karim Chaudhry. "Optimization of material delivery time analysis by using Visual Basic for applications in Excel." *Journal of Applied Research in Technology & Engineering* 2, no. 2 (2021): 89-100.

[51] Kalwar, Muhammad Ahmed, Muhammad Ali Khan, Muhammad Faisal Shahzad, Muzamil Hussain Wadho, and Hussain Bux Marri. "Development of linear programming model for optimization of product mix and maximization of profit: case of leather industry." *Journal of Applied Research in Technology & Engineering* 3, no. 1 (2022): 67-78.

[52] Kalwar, Muhammad Ahmed, Muhammad Faisal Shahzad, Muzamil Hussain Wadho, Muhammad Ali Khan, and Shakeel Ahmed Shaikh. "Automation of order costing analysis by using Visual Basic for applications in Microsoft Excel." *Journal of Applied Research in Technology & Engineering* 3, no. 1 (2022): 29-59.

[53] Chaudhry, Abdul Karim, Muhammad Ahmed Kalwar, Muhammad Ali Khan, and Shakeel Ahmed Shaikh. "Improving the Efficiency of Small Management Information System by Using VBA." *International Journal of Science and Engineering Investigations* 10, no. 111 (2021): 7-13.

[54] Kalwar, Muhammad Ahmed, Asif Nawaz Wassan, Zuhaib Phul, Muzamil Hussain Wadho, Tanveer Sarwar Malik, and Muhammad Ali Khan. "Automation of material cost comparative analysis report using VBA Excel: a case of footwear company of Lahore." *Journal of Applied Research in Technology & Engineering* 4, no. 1 (2023): 13-23.

[55] Khan, Muhammad Ali, Sarmad Ali Khaskheli, Hamid Ali Kalwar, Muhammad Ahmed Kalwar, Hussain Bux Marri, and Murlidhar Nebhwani. "Improving the Performance of Reception and OPD by Using Multi-Server Queuing Model in Covid-19 Pandemic." *International Journal of Science and Engineering Investigations* 10, no. 113 (2021): 20-29.

[56] Khaskheli, Sarmad Ali, Hamid Ali Kalwar, Muhammad Ahmed Kalwar, Hussain Bux Marri, Muhammad Ali Khan, and Murlidhar Nebhwani. "APPLICATION OF MULTI-SERVER QUEUING MODEL TO ANALYZETHE QUEUING SYSTEM OF OPD DURING COVOD-19 PANDEMIC: A CASE STUDY." *Journal of Contemporary Issues in Business and Government Vol* 27, no. 05 (2021).

[57] Haines, I. E., and M. P. Jones. "When a system breaks: a queuing theory model for the number of intensive care beds needed during the COVID-19 pandemic." *Med. J. Aust* (2020).

[58] Meares, Hamish DD, and Michael P. Jones. "When a System Breaks: A Queuing Theory Model for the Number of Intensive Intensive Intensive Care Beds Needed Dur-ing the COVID-19 Pandemic."

[59] Mittal, Himanshu, and Naresh Sharma. "A probabilistic model for the assessment of queuing time of coronavirus disease (COVID-19) patients using queuing model." *Technology* 11, no. 8 (2020): 22-31.

[60] Zimmerman, Samantha L., Alexander R. Rutherford, Alexa van der Waall, Monica Norena, and Peter Dodek. "A queuing model for ventilator capacity management during the COVID-19 pandemic." *Health Care Management Science* (2023): 1-17.

[61] Kalwar, Muhammad Ahmed, Hussain Bux Marri, Muhammad Ali Khan, and Sarmad Ali Khaskheli. "Applications of queuing theory and discrete event simulation in health care units of Pakistan." *International Journal of Science and Engineering Investigations* 10, no. 109 (2021): 6-18.

[62] Khaskheli, Sarmad Ali, Hussain Bux Marri, Murlidhar Nebhwani, Muhammad Ali Khan, and Muhammad Ahmed. "Compartive Study of Queuing Systems of Medical Out Patient Departments of Two Public Hospitals." In *Proceedings of the International Conference on Industrial Engineering and Operations Management*, vol. 1913, pp. 2702-2720. 2020.

[63] Cirrincione, Luigi, Fulvio Plescia, Caterina Ledda, Venerando Rapisarda, Daniela Martorana, Raluca Emilia Moldovan, Kelly Theodoridou, and Emanuele Cannizzaro. "COVID-19 pandemic: Prevention and protection measures to be adopted at the workplace." *Sustainability* 12, no. 9 (2020): 3603.

[64] Abualigah, Laith Mohammad, Ahamad Tajudin Khader, and Essam Said Hanandeh. "A hybrid strategy for krill herd algorithm with harmony search algorithm to improve the data clustering**1**." *Intelligent Decision Technologies* 12, no. 1 (2018): 3-14.

[65] Raskutti, Garvesh, Martin J. Wainwright, and Bin Yu. "Early stopping and non-parametric regression: an optimal data-dependent stopping rule." *The Journal of Machine Learning Research* 15, no. 1 (2014): 335-366.

[66] Franceschi, Jean-Yves, Edouard Delasalles, Mickaël Chen, Sylvain Lamprier, and Patrick Gallinari. "Stochastic latent residual video prediction." In *International Conference on Machine Learning*, pp. 3233-3246. PMLR, 2020.

[67] Kalra, Kanav. "Predicting the Genotype to Phenotype relationship in Plants using Machine Learning and Deep Learning." PhD diss., University of Saskatchewan, 2023.

[68] Ozery-Flato, Michal, Liat Ein-Dor, Ora Pinchasov, Miel Dabush Kasa, Efrat Hexter, Gabriel Chodick, Michal Rosen-Zvi, and Michal Guindy. "Impact of the COVID-19 Pandemic on Clinical Findings in Medical Imaging Exams in a Nationwide Israeli Health Organization: Observational Study." *JMIR Formative Research* 7, no. 1 (2023): e42930.

[69] Data Gov, "COVID-19 Data Set," COVID-19 Repository, Ministry of Health, Government of Israel, Isarael, 2020. https://data.gov.il/dataset/covid-19/resource/3f5c975e-7196-454b-8c5b-ef85881f78db/download/-readme.pdf

[70] Dey, Priyadarshini, S. Vaijayanthimala, Vivek S. Dalvi, Ashu Jain, Deepak Gola, Medha Bajpai, Randhir K. Bharti, and Nitin Chauhan. "COVID-19: understanding the pandemic emergence, impact and infection prevalence worldwide." *J. Pure Appl. Microbiol* 14 (2020): 2235-2251.

[71] Piccialli, Francesco, Vincenzo Schiano Di Cola, Fabio Giampaolo, and Salvatore Cuomo. "The role of artificial intelligence in fighting the COVID-19 pandemic." *Information Systems Frontiers* 23, no. 6 (2021): 1467-1497.

[72] Doyle, Ronan, and Kieran Conboy. "The role of IS in the covid-19 pandemic: A liquid-modern perspective." *International Journal of Information Management* 55 (2020): 102184.

[73] Heywood, Anita E., and C. Raina Macintyre. "Elimination of COVID-19: what would it look like and is it possible?." *The Lancet Infectious Diseases* 20, no. 9 (2020): 1005-1007.

[74] Chung, Sheng-Chia, Sushila Marlow, Nicholas Tobias, Alessio Alogna, Ivano Alogna, San-Lin You, Kamlesh Khunti, Martin McKee, Susan Michie, and Deenan Pillay. "Lessons from countries implementing find, test, trace, isolation and support policies in the rapid response of the COVID-19 pandemic: a systematic review." *BMJ open* 11, no. 7 (2021): e047832.

[75] Stawicki, Stanislaw P., Rebecca Jeanmonod, Andrew C. Miller, Lorenzo Paladino, David F. Gaieski, Anna Q. Yaffee, Annelies De Wulf et al. "The 2019–2020 novel coronavirus (severe acute respiratory syndrome coronavirus 2) pandemic: A joint american college of academic international medicine-world academic council of emergency medicine multidisciplinary COVID-19 working group consensus paper." *Journal of global infectious diseases* 12, no. 2 (2020): 47.

[76] Dong, Di, Zhenchao Tang, Shuo Wang, Hui Hui, Lixin Gong, Yao Lu, Zhong Xue et al. "The role of imaging in the detection and management of COVID-19: a review." *IEEE reviews in biomedical engineering* 14 (2020): 16-29.

[77] Aljondi, Rowa, and Salem Alghamdi. "Diagnostic value of imaging modalities for COVID-19: scoping review." *Journal of medical Internet research* 22, no. 8 (2020): e19673.

[78] Huang, Zixing, Shuang Zhao, Zhenlin Li, Weixia Chen, Lihong Zhao, Lipeng Deng, and Bin Song. "The battle against coronavirus disease 2019 (COVID-19): emergency management and infection control in a radiology department." *Journal of the american college of radiology* 17, no. 6 (2020): 710-716.

[79] Desai, Sudhen B., Anuj Pareek, and Matthew P. Lungren. "Deep learning and its role in COVID-19 medical imaging." *Intelligence-Based Medicine* 3 (2020): 100013.

[80] Song, Jiangdian, Hongmei Wang, Yuchan Liu, Wenqing Wu, Gang Dai, Zongshan Wu, Puhe Zhu, Wei Zhang, Kristen W. Yeom, and Kexue Deng. "End-to-end automatic differentiation of the coronavirus disease 2019 (COVID-19) from viral pneumonia based on chest CT." *European journal of nuclear medicine and molecular imaging* 47 (2020): 2516-2524.

[81] Chitra, R., and V. Seenivasagam. "Review of heart disease prediction system using data mining and hybrid intelligent techniques." *ICTACT journal on soft computing* 3, no. 04 (2013): 605-09.

[82] Anooj, P. K. "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules." *Journal of King Saud University-Computer and Information Sciences* 24, no. 1 (2012): 27-40.

[83] Ge, Yanqiu, Qinghua Wang, Li Wang, Honghu Wu, Chen Peng, Jiajing Wang, Yuan Xu, Gang Xiong, Yaoyun Zhang, and Yingping Yi. "Predicting post-stroke pneumonia using deep neural network approaches." *International journal of medical informatics* 132 (2019): 103986.

[84] Hooda, Rahul, Ajay Mittal, and Sanjeev Sofat. "An efficient variant of fully-convolutional network for segmenting lung fields from chest radiographs." *Wireless Personal Communications* 101 (2018): 1559-1579.

[85] Rajaraman, Sivaramakrishnan, Sema Candemir, Incheol Kim, George Thoma, and Sameer Antani. "Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs." *Applied Sciences* 8, no. 10 (2018): 1715.

[86] Faisal, Muhammad, Andrew J. Scally, Natalie Jackson, Donald Richardson, Kevin Beatson, Robin Howes, Kevin Speed et al. "Development and validation of a novel computer-aided score to predict the risk of in-hospital mortality for acutely ill medical admissions in two acute hospitals using their first electronically recorded blood test results and vital signs: a cross-sectional study." *BMJ open* 8, no. 12 (2018): e022939.

[87] Faisal, Muhammad, Andy Scally, Donald Richardson, Kevin Beatson, Robin Howes, Kevin Speed, and Mohammed A. Mohammed. "Development and external validation of an automated computer-aided risk score for predicting sepsis in emergency medical admissions using the patient's first electronically recorded vital signs and blood test results." *Critical care medicine* 46, no. 4 (2018): 612-618.

[88] Atlam, Mostafa, Hanaa Torkey, Nawal El-Fishawy, and Hanaa Salem. "Coronavirus disease 2019 (COVID-19): Survival analysis using deep learning and Cox regression model." *Pattern Analysis and Applications* 24 (2021): 993-1005.

[89] Tiwari, Dimple, Bhoopesh Singh Bhati, Fadi Al-Turjman, and Bharti Nagpal. "Pandemic coronavirus disease (Covid-19): World effects analysis and prediction using machine-learning techniques." *Expert Systems* 39, no. 3 (2022): e12714.

[90] Jiang, Xiangao, Megan Coffee, Anasse Bari, Junzhang Wang, Xinyue Jiang, Jianping Huang, Jichan Shi et al. "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity." *Computers, Materials & Continua* 63, no. 1 (2020): 537-551.

[91] Khan, Muzammil, Muhammad Taqi Mehran, Zeeshan Ul Haq, Zahid Ullah, Salman Raza Naqvi, Mehreen Ihsan, and Haider Abbass. "Applications of artificial intelligence in COVID-19 pandemic: A comprehensive review." *Expert systems with applications* 185 (2021): 115695.

[92] Islam, Muhammad Nazrul, Toki Tahmid Inan, Suzzana Rafi, Syeda Sabrina Akter, Iqbal H. Sarker, and AKM Najmul Islam. "A systematic review on the use of AI and ML for fighting the COVID-19 pandemic." *IEEE Transactions on Artificial Intelligence* 1, no. 3 (2020): 258-270.

[93] Souza Jr, José Leão de, Vanessa Damazio Teich, Anna Carolina Batista Dantas, Daniel Tavares Malheiro, Marcio Aparecido de Oliveira, Eduardo Segalla de Mello, and Miguel Cendoroglo Neto. "Impact of the COVID-19 pandemic on emergency department visits: experience of a Brazilian reference center." *Einstein (Sao Paulo)* 19 (2021).

[94] Schwab, Patrick, August DuMont Schütte, Benedikt Dietz, and Stefan Bauer. "Clinical predictive models for COVID-19: systematic study." *Journal of medical Internet research* 22, no. 10 (2020): e21439.

[95] Purdie, Anna, Sarah Hawkes, Kent Buse, Kristine Onarheim, Wafa Aftab, Nicola Low, and Sonja Tanaka. "Sex, gender and COVID-19: disaggregated data and health disparities." *BMJ Global Health Blog* 24 (2020): 2020.

[96] Whittington, A. M., S. Logan, A. Goodman, A. Houston, L. John, and F. Wrigley. "Coronavirus: rolling out community testing for COVID-19 in the NHS." *BMJ opinion* (2020).

[97] Menni, Cristina, Ana M. Valdes, Maxim B. Freidin, Carole H. Sudre, Long H. Nguyen, David A. Drew, Sajaysurya Ganesh et al. "Real-time tracking of self-reported symptoms to predict potential COVID-19." *Nature medicine* 26, no. 7 (2020): 1037-1040.

[98] Menni, Cristina, Ana M. Valdes, Maxim B. Freidin, Sajaysurya Ganesh, Julia S. El-Sayed Moustafa, Alessia Visconti, Pirro Hysi et al. "Loss of smell and taste in combination with other symptoms is a strong predictor of COVID-19 infection." *MedRxiv* (2020): 2020-04.

[99] Guleria, Pratiyush, Shakeel Ahmed, Abdulaziz Alhumam, and Parvathaneni Naga Srinivasu. "Empirical study on classifiers for earlier prediction of COVID-19 infection cure and death rate in the Indian states." In *Healthcare*, vol. 10, no. 1, p. 85. MDPI, 2022.

[100] Hu, Chuanyu, Zhenqiu Liu, Yanfeng Jiang, Oumin Shi, Xin Zhang, Kelin Xu, Chen Suo et al. "Early prediction of mortality risk among patients with severe COVID-19, using machine learning." International journal of epidemiology 49, no. 6 (2020): 1918-1929.

[101] John Lu, Z. Q. "The elements of statistical learning: data mining, inference, and prediction." (2010): 693-694.

[102] Ruppert, David. "The elements of statistical learning: data mining, inference, and prediction." (2004): 567-567.

[103] Hastie, Trevor, Robert Tibshirani, Jerome H. Friedman, and Jerome H. Friedman. The elements of statistical learning: data mining, inference, and prediction. Vol. 2. New York: springer, 2009.

[104] Fernández-Delgado, Manuel, Eva Cernadas, Senén Barro, and Dinani Amorim. "Do we need hundreds of classifiers to solve real world classification problems?." *The journal of machine learning research* 15, no. 1 (2014): 3133-3181.

[105] Wynants, Laure, Ben Van Calster, Gary S. Collins, Richard D. Riley, Georg Heinze, Ewoud Schuit, Elena Albu et al. "Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal." *bmj* 369 (2020).

[106] Zoabi, Yazeed, and Noam Shomron. "COVID-19 diagnosis prediction by symptoms of tested individuals: a machine learning approach." *MedRxiv* (2020): 2020-05.

[107] Zoabi, Yazeed, Orli Kehat, Dan Lahav, Ahuva Weiss-Meilik, Amos Adler, and Noam Shomron. "Predicting bloodstream infection outcome using machine learning." *Scientific reports* 11, no. 1 (2021): 20101.

[108] Artzi, Nitzan Shalom, Smadar Shilo, Eran Hadar, Hagai Rossman, Shiri Barbash-Hazan, Avi Ben-Haroush, Ran D. Balicer, Becca Feldman, Arnon Wiznitzer, and Eran Segal. "Prediction of gestational diabetes based on nationwide electronic health records." *Nature medicine* 26, no. 1 (2020): 71-76.

[109] Abou Omar, Kamil Belkhayat. "XGBoost and LGBM for Porto Seguro's Kaggle challenge: A comparison." *Preprint Semester Project* (2018).

[110] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785-794. 2016.

[111] Park, Min Sue, Hyeontae Jo, Haeun Lee, Se Young Jung, and Hyung Ju Hwang. "Machine learning-based COVID-19 patients triage algorithm using patient-generated health data from nationwide multicenter database." *Infectious Diseases and Therapy* 11, no. 2 (2022): 787-805.

[112] Vieira, Bruno Hebling, Franziskus Liem, Kamalaker Dadi, Denis A. Engemann, Alexandre Gramfort, Pierre Bellec, Richard Cameron Craddock et al. "Predicting future cognitive decline from non-brain and multimodal brain imaging data in healthy and pathological aging." *Neurobiology of aging* 118 (2022): 55-65.

[113] Jaiswal, Sunit Kumar, and Sudhir Agarwal. "Covid 19 diagnosis prediction using Machine Learning."

[114] Josse, Julie, Nicolas Prost, Erwan Scornet, and Gaël Varoquaux. "On the consistency of supervised learning with missing values." *arXiv preprint arXiv:1902.06931* (2019).

[115] Kim, Kakyeong, Yoonjung Yoonie Joo, Gun Ahn, Hee-Hwan Wang, Seo-Yoon Moon, Hyeonjin Kim, Woo-Young Ahn, and Jiook Cha. "The sexual brain, genes, and cognition: A machine-predicted brain sex score explains individual differences in cognitive intelligence and genetic influence in young children." *Human Brain Mapping* 43, no. 12 (2022): 3857-3872.

[116] Burkhardt, Daniel B., Jay S. Stanley III, Alexander Tong, Ana Luisa Perdigoto, Scott A. Gigante, Kevan C. Herold, Guy Wolf, Antonio J. Giraldez, David van Dijk, and Smita Krishnaswamy. "Quantifying the effect of experimental perturbations at single-cell resolution." *Nature biotechnology* 39, no. 5 (2021): 619-629.

[117] Silverman, Bernard W. "Density estimation for statistics and data analysis". Vol. 26. CRC press, 1986.

[118] Kung, Yi-Hung, Pei-Sheng Lin, and Cheng-Hsiung Kao. "An optimal k-nearest neighbor for density estimation." Statistics & Probability Letters 82, no. 10 (2012): 1786-1791.

[119] Biau, Gérard, Frédéric Chazal, David Cohen-Steiner, Luc Devroye, and Carlos Rodriguez. "A weighted k-nearest neighbor density estimate for geometric inference." (2011): 204-237.

[120] Mack, Y. P., and Murray Rosenblatt. "Multivariate k-nearest neighbor density estimates." *Journal of Multivariate Analysis* 9, no. 1 (1979): 1-15.

[121] Baig, Mirza Azam, Sarmad Ahmed Shaikh, Kamlesh Kumar Khatri, Muneer Ahmed Shaikh, Muhammad Zohaib Khan, and Mahira Abdul Rauf. "Prediction of Students Performance Level Using Integrated Approach of ML Algorithms." *International Journal of Emerging Technologies in Learning* 18, no. 1 (2023).

[122] Channar, Paras Bashir, Ghalib Ahmed, Jalil Ahmed Thebo, Muhammad Ali Khan, and Liaquat Ali Rahoo. "Factors Of Knowledge Sharing Among Faculty Members In Higher Educational Institutions: An Empirical Study Of The Public Sector." *Journal of Positive School Psychology* (2023): 1498-1506.

[123] Rahoo, L. A., M. A. Khan, M. A. Buriro, Z. H. Baladi, and M. S. Abbasi. "Evaluation of Information Services from the Perspective of Faculties and Evaluation of Information Services from the Perspective of Faculties and Students of Mehran University Engineering and Technology, Jamshoro Pakistan." *International Journal of Disaster Recovery and Business Continuity* 11, no. 1 (2020): 1526-1538.

[124] Rahoo, Liaquat Ali, Muhammad Ali Khan Nagar, and Arabella Bhutto. "The Use of Information Retrieval Tools by the Postgraduate Students of Higher Educational Institutes of Pakistan." *Asian Journal of contemporary education* 3, no. 1 (2019): 59-64.

[125] Buriro, Mumtaz Ali, Liaquat Ali Rahoo, Muhammad Ali Khan Nagar, Maryam Kalhoro, Shadab Kalhoro, and Azam Ali Halepota. "Social Media used for promoting the Libraries and Information Resources and services at University Libraries of Sindh Province." In *2018 IEEE international conference on innovative research and development (ICIRD)*, pp. 1-4. IEEE, 2018.

[126] Gul, Mehmood Ali, Zartashia Kiran Imran, Muhammad Ahmed Kalwar, Muzamil Hussain Wadho, Muhammad Ali Khan, Liaquat Ali Raho, and Shakeel Ahmed Shaikh. "The perception of online banking customers about internet services on its cost effectiveness, ease of use, easily availability and efficiency & effectiveness", International Research Journal of Modernization in Engineering Technology and Science, Volume:04/Issue:11/November-2022.

[127] Nagar, Muhammad Ali Khan, Liaquat Ali Rahoo, Hafiz Abdur Rehman, and Sadia Arshad. "Education management information systems in the primary schools of sindh a case study of hyderabad division." In *2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, pp. 1-5. IEEE, 2018.

[128] Ay, Şevket, Ekin Ekinci, and Zeynep Garip. "A comparative analysis of meta-heuristic optimization algorithms for feature selection on ML-based classification of heart-related diseases." *The Journal of Supercomputing* (2023): 1-30.

[129] Kaur, Rajvir. "A comparative analysis of selected set of natural language processing (NLP) and machine learning (ML) algorithms for clinical coding using clinical classification standards." PhD diss., Western Sydney University (Australia), 2018.

[130] De Souza, Bruno Feres, André CPLF de Carvalho, and Carlos Soares. "A comprehensive comparison of ml algorithms for gene expression data classification." In *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8. IEEE, 2010.

[131] Tanriver, Gizem, Merva Soluk Tekkesin, and Onur Ergen. "Automated detection and classification of oral lesions using deep learning to detect oral potentially malignant disorders." *Cancers* 13, no. 11 (2021): 2766.

[132] Welikala, Roshan Alex, Paolo Remagnino, Jian Han Lim, Chee Seng Chan, Senthilmani Rajendran, Thomas George Kallarakkal, Rosnah Binti Zain et al. "Automated detection and classification of oral lesions using deep learning for early detection of oral cancer." *IEEE Access* 8 (2020): 132677-132693.

[133] Yandex, "Classification of Algorithms in Machine Learning," 2022.

[134] Takhtardeshir, Soheib, Mohammad Mahdi Moradi, and Reza Ghaderi. "Automated detection of COVID-19 cases through Transfer Learning using X-ray Imaging Data." (2021).

[135] Ishii, Euma, Nobutoshi Nawa, Satoru Hashimoto, Hidenobu Shigemitsu, and Takeo Fujiwara. "Development, validation, and feature extraction of a deep learning model predicting in-hospital mortality using Japan's largest national ICU database: a validation framework for transparent clinical Artificial Intelligence (cAI) development." *Anaesthesia Critical Care & Pain Medicine* 42, no. 2 (2023): 101167.

[136] Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." *Advances in neural information processing systems* 30 (2017).

[137] Lundberg, Scott M., Bala Nair, Monica S. Vavilala, Mayumi Horibe, Michael J. Eisses, Trevor Adams, David E. Liston et al. "Explainable machine-learning predictions for the prevention of hypoxaemia during surgery." *Nature biomedical engineering* 2, no. 10 (2018): 749-760.

[138] Elhia, Mohamed, Khalid Chokri, and Meryem Alkama. "Optimal control and free optimal time problem for a COVID-19 model with saturated vaccination function." *Commun. Math. Biol. Neurosci.* 2021 (2021): Article-ID.

[139] Liang, Shuang, Huixiang Liu, Yu Gu, Xiuhua Guo, Hongjun Li, Li Li, Zhiyuan Wu, Mengyang Liu, and Lixin Tao. "Fast automated detection of COVID-19 from medical images using convolutional neural networks." *Communications Biology* 4, no. 1 (2021): 35.

[140] Johnson, Roger W. "An introduction to the bootstrap." *Teaching statistics* 23, no. 2 (2001): 49-54.

[141] Chernick, Michael R., and Robert A. LaBudde. *An introduction to bootstrap methods with applications to R*. John Wiley & Sons, 2014.

[142] Zoabi, Yazeed, Shira Deri-Rozov, and Noam Shomron. "Machine learning-based prediction of COVID-19 diagnosis based on symptoms." *npj digital medicine* 4, no. 1 (2021): 3.

[143] Voinsky, Irena, Gabriele Baristaite, and David Gurwitz. "Effects of age and sex on recovery from COVID-19: Analysis of 5769 Israeli patients." *Journal of Infection* 81, no. 2 (2020): e102-e103.

[144] Kuodi, Paul, Yanay Gorelik, Hiba Zayyad, Ofir Wertheim, Karine Beiruti Wiegler, Kamal Abu Jabal, Amiel A. Dror, Saleh Nazzal, Daniel Glikman, and Michael Edelstein. "Association between BNT162b2 vaccination and reported incidence of post-COVID-19 symptoms: cross-sectional study 2020-21, Israel." *npj Vaccines* 7, no. 1 (2022): 101.

[145] Caspi, Gil, Avshalom Dayan, Yael Eshal, Sigal Liverant-Taub, Gilad Twig, Uri Shalit, Yair Lewis, Avi Shina, and Oren Caspi. "Socioeconomic disparities and COVID-19 vaccination acceptance: a nationwide ecologic study." *Clinical Microbiology and Infection* 27, no. 10 (2021): 1502-1506.

[146] Arshed, Muhammad Asad, Wajeeha Qureshi, Muhammad Usman Ghani Khan, and Muhammad Abdul Jabbar. "Symptoms based Covid-19 disease diagnosis using machine learning approach." In *2021 International Conference on Innovative Computing (ICIC)*, pp. 1-7. IEEE, 2021.

[147] Dabbura, Imad. "K-means clustering: Algorithm, applications, evaluation methods, and drawbacks." *Towards Data Science* (2018).

[148] Simón Sánchez, Alejandro-Martín, José González-Piqueras, Luis de la Ossa, and Alfonso Calera. "Convolutional neural networks for agricultural land use classification from Sentinel-2 image time series." *Remote Sensing* 14, no. 21 (2022): 5373.

[149] Nicholson, Chris. "A Beginner's Guide to Multilayer Perceptrons (MLP), Pathmind." *Accessed on* 8 (2020): 16.

[150] Chauhan, N. S. "Decision tree algorithm, explained—Kdnuggets." (2020).