Covid-19 Disease Prediction System

Prabhat Jindal[1], Taruna[2]

[1]Dept of Computer Science & Engineering
[2]Assistant Professor, Dept of Computer Science & Engineering
[1, 2] International Institute of Technology & management, Sonepat, Haryana, INDIA

Abstract- *The focus of this paper is to determine the functionalities of different data mining techniques that can be used to find whether a person is suffering from Corona Virus Disease or not. Data Mining has proven its use in the medical world for all different types of diseases. There are many different techniques available that can be very helpful is developing a model to create a sophisticated way of pre-examining a patient having covid-19 symptoms or not. We will be talking more on this in the paper.*

Keywords*- Prediction,COVID-19,Technology*

I. INTRODUCTION

The disease prediction system only provides possible results but does not guarantee that it predicts the disease correctly or perfectly. But it has significantly higher accuracy for predicting possible diseases. In our duration of research, we have analysed the accuracy of this system for 2 different diseases and our accuracy can be as high as 89%. The aim of this project is to verify if a patient is likely to be diagnosed with any cardiovascular heart disease based on their medical attributes like gender, age, chest pain, fasting blood sugar etc. A public database describes a data repository used for research and dedicated to housing data related to scientific research on an open platform. These databases collect and store heterogeneous and multidimensional health, medical and scientific research in a structured form and characteristics of crowd/multi-ownership, complexity and security. These databases cover a wide range of data, including those related to cancer disease research, Covid research, burden of disease, nutrition and health, and genetics and the environment. Researchers may request access to data based on the scope of the database and application procedures needed to conduct the relevant medical research. In India, the first case of COVID-19 was reported on 30 January,2020. By 17 July, 2020 there were 1,003,832 confirmed cases of COVID-19, of which 373,379 were active cases, In India, 677,422 cured/discharged, 1 migrated and 26,816 deaths were reported. Diagram shows the COVID-19 epidemic broke out in India (Hollingsworth et al. 2020). Compared to global death rate 6.19%, death rate in India is relatively low at 2.83%5. India has the highest number of confirmed cases in Asia.

WHAT IS CORONA VIRUS?

Coronaviruses are a large family of enveloped & non-segmented & single-stranded & positive sense RNA viruses that circulate among animals including camels & cats & bats. Coronaviruses derive their name from their electron microscopic image which does resembles a crown – or corona. COVID-19 (coronavirus disease 2019) is a disease caused by the SARS-CoV-2 virus and was discovered in December 2019 in Wuhan city, China. It is very highly contagious and has spread rapidly throughout the world.

COVID-19 most commonly causes respiratory symptoms that can resemble a cold & flu & pneumonia. COVID-19 can attack more than your lungs and respiratory system. Many other parts of your body can also be affected by the disease.

- Max people with COVID-19 have mild symptoms, but some people become seriously ill.
- Some people, including those with minimum or no symptoms, may suffer from postCOVID conditions, or "long COVID".
- Older population who have certain underlying medical conditions are at increased risk of serious illness from COVID-19.

Like many other respiratory viruses, the corona virus spread quickly through droplets that you spray out of your mouth or nose when you breathe & cough & sneeze or talk. The word corona means crown and refers to the appearance that coronaviruses get from the peak proteins that stick out of them. These spike proteins are important for the biology of this virus. The protein spike is the part of the virus that binds to a human cell to infect it & allowing it to replicate inside the cell and spread to other cells.

Developmnt of some antibodies can protect you from SARS-CoV-2 by targeting these peak proteins. Because of the importance of this specific part of the virus & scientists who sequence the virus for research are constantly watching for mutations that cause changes in the spike protein through a process called genomic surveillance. As the genetic changes of the virus occur over period of time, the SARS-CoV-2 virus

begins to form genetic lineages. Just like a family has a family tree the SARS-CoV-2 virus can be similarly plotted. Sometimes the branches of this tree have different attributes that change how quickly the virus spreads or the severity of the disease it causes or the effectiveness of treatments against it. Scientists has named these viruses with these changes as "variants." They are still SARS-CoV-2, but they may behave differently.

DATA MINING APPROACH

Data mining is a multidisciplinary field at the border of database technologies, statistics, ML and pattern recognition, who are professionals in all these fields. Although this approach is not yet widespread in medical research, several studies have shown the promise of data mining in creating disease prediction models & assessing patient risk & helping physicians make clinical decisions.

DATA-MINING MODELS

1. Predictive Model: A predictive model represents prediction values about data using known results found from different data. Predictive modelling can be done based on the use of variant historical data. Data mining predictive model tasks include regression & time series analysis & classification and prediction. A predictive model is known as statistical regression model. It is learning & monitoring technique that involves explaining the dependence of several attribute values on the values of other attributes in a similar item and growing a model that can predict those attribute values for recent instances.

2. Description Model: A descriptive model distinguishes relationships or patterns in data. Unlike a predictive model, a descriptive model serves as a way to explore the properties of the data under investigation, not to predict new properties & clustering & summarization & association rules, and sequence discovery are tasks of descriptive data modelling. Descriptive analytics Focus on summarizing and converting data into important information for monitoring and reporting.

DATA MINING TASKS

A model is typically implemented by a task, with the objective of the description being to generalize the patterns of potential associations in the data. Therefore, using a definitive model usually results in several collections with the same or similar attributes. Prediction mainly refers to the estimation of the variable value of a particular attribute based on the variable values of other attributes that includes classification & regression.

DATA MINING ALGORITHMS FOR CLINICALLY HUGE DATA

Data mining based on clinical big data can produce effective and valuable knowledge that is essential for accurate clinical decision making and risk assessment. Data mining algorithms enable the realization of these goals.

Supervised learning: A concept often mentioned in supervised learning is partitioning data sets. To avoid overfitting the model, the data set can generally be divided into two or three parts: the training set & the validation set & last the test set. Unsupervised learning: In many data analysis processes and the amount of usable identified data is small & data identification is a lengthy process. Unsupervised learning is essential to assess and categorize data according to similarities & characteristics& correlations & has three main applications: 1. data clustering, 2. association analysis, and last dimensionality reduction. Hence, the unsupervised learning method presented in this section include cluster analysis & association rules & and PCA.

SVMs: SVM is altogether a new classification & prediction method developed by Cortes and Vapnik and is a data-driven approach that does not require assumptions about the distribution of the data. The main purpose of SVM is to identify a separation boundary (called a hyperplane) to help classify cases; thus, the advantages of SVM are obvious when classifying and predicting cases based on data with a large dimension or data with a small sample size.

Competitive risk model: Kaplan-Meier marginal regression and the Cox proportional hazards model are widely used in survival analysis in clinical trials. Classical survival analysis usually considers only one endpoint, such as the effect of patient survival time. However, Using clinical medical research & multiple endpoints usually coexist and these endpoints compete with each other to generate competing risk data. In the case of multiple observed events, the use of a single endpoint analysis method may lead to a biased estimate of the probability of the endpoints due to the existence of competing risks. The competing risk model is a classical statistical model based on the data distribution hypothesis. Its main advantage is its accurate estimation of the cumulative incidence of outcomes for right-censored survival data with multiple endpoints. In data analysis, the cumulative hazard rate is estimated using the cumulative incidence function in univariate analysis, and Gray's test is used for comparisons between groups.

DATA MINING ARCHITECTURE

The most important components of data mining systems are a data source & data mining engine, data warehouse server and the pattern evaluation module & graphical user interface and knowledge base.

Data Source: The real data source is database, data warehouse, World Wide Web (WWW), text files and other documents. For data mining to be successful, you need a huge amount of historical data. Organizations typically store data in databases or data warehouses. Data warehouses can contain one or more databases, word processors, or other data stores. Sometimes the information can also contain plain text files or tables. Another primary source of data is the World Wide Web or the Internet.

Different processes: Data needs to be cleaned & integrated & selected before being passed to a database or data warehouse server. Since the information comes from different sources and in different formats, it cannot be directly used for the data mining procedure because the data may not be complete and accurate. Therefore, the first data must be cleaned and précised. It provides more information than necessary will be gathered from various data sources and only the data of interest will need to be selected and passed to the server. These procedures are not as simple as we think. Several methods can be performed on the data for selection, integration, and cleaning. Database or Data Warehouse Server: A database or data warehouse server consists of raw data that is ready to be processed. So the server is the reason for getting the relevant data which is based on the data mining according to the user request. Data Mining Engine: A data mining engine is a major part of any data mining system. It contains multiple modules for performing data mining tasks and including association & characterization & classification and clustering and prediction and last time line analysis. In other words, data mining is the base of our data mining architecture. It includes tools and software used to derive insights and knowledge from data collected from various data sources and stored in a data warehouse.

Pattern Evaluation Module: The pattern evaluation module is primarily responsible for the extent to which a pattern is examined using a threshold value. It works with a data mining engine to focus your search on exciting patterns. Bet measurement is commonly used in this segment, which works with data mining engines to focus the search on fascinating patterns. It can use a bet threshold to filter out discovered similarities. On the other side, the pattern evaluation module can be coordinated with the mining module depending on the implementation of the data mining techniques used. For effective data mining, it is abnormally recommended to push the evaluation of pattern stakes as far into the mining process

as possible to limit the search to only fascinating patterns. Graphical User Interface: The graphical user interface (GUI) module role is to communicates between the data mining system and the end user. This module helps the user to use the system easily and effectively without knowing the complexities of the process. This module communicates with the data mining system when the user enters a query or task and displays the results for some input.

Knowledge Base: The knowledge base is helpful in the entire data mining process. It can be useful to conduct a search or evaluate a bet of result patterns. The knowledge base can even contain user insights 9 and user experience data that can be useful in the data mining process. The data mining engine can take request from the knowledge base database to make the result more accurate and reliable and trustworthy. The pattern evaluation engine regularly interacts with the knowledge base to obtain input and also to update it.

II. LITERATURE REVIEW

Scientists and researchers have long viewed the value of sound as its ability to predict behavior and health. For example, digital stethoscopes use special external microphone recorders to check heart or lung sounds. They also require a doctor trained in listening and interpretation and have recently been replaced by a variety of imaging techniques (such as MRI, ultrasonography) that are easy to analyze and interpret. However, recent changes in audio interpretation and modeling have the potential to reverse these trends and provide audio as a cheap and easy delivery alternative. Recently, microphones have been used for sound processing in products such as smartphones and wearables.

Wang Yunlu et al proposed different classification method for large-scale surveillance of people infected with COVID-19; this work can be used to identify various breathing patterns and we can apply this tool to the real world. This article first introduces a new powerful RS (Respiratory Simulation) model to extract the difference between training data and the lack of real-world data to reflect the characteristics of real breathing. They first used two neural network methods such as the GRU network monitoring tool (BI_at_GRU) (Tachic, Eupnoea, Biotas, Cheyne-Stokes, Bradypnea, and Central-Apnea) to identify six important respiratory patterns in the hospital. The results showed that 6 different respiratory rates were 94.5%, 94.4%, 95.1% and 94%. The model can determine accuracy, precision, recall, and F1 by 8% each. In comparative studies, the BI_at_GRU obtained for classification of breath of air outperforms existing models. Its deep design concept and design has great potential

for large-scale uses such as sleeping quarters, public spaces and workplaces.

In 2013, Jiang Cheng et al. proposed a portable, contactless system. Monitor the health of people wearing masks by analyzing respiratory characteristics. The device has only thermal camera with FLIR (Forward Looking Infrared) function and Android device. In real-world situations, such as pre-screening of hospitals and medical facilities, this can help differentiate different COVID-19 patients. In this work, they combine thermal and RGB video from a camera based on DL architecture for diagnostics. First, they used a lung analysis technique to identify people wearing masks; they used the BI_at_GRU function for results of lung diseases to obtain health screening and finally classified patients' respiratory diseases with 83.7% accuracy. Chapter Han Jing and others. prepared a study on intelligent analysis of COVID-19 speech data considering four factors: i. sleep well, ii.Weight three. Worry, fourth. burnout. Han Jing and others. The data was collected by the "COVID-19 Voice App" by researchers and researchers at the University of Cambridge and the "Corona Voice Detection App" by researchers at Mellon University.

After processing the data, these individuals received a total of 378 partitions; Based on this preliminary study, they collected 260 records for the future. Collected 256 audio clips from 50 COVID-19 patients; for future work, switch to a range with a sample rate of 0.016 MHz. They considered two noise sources in this study, CompParE and eGeMAPS; both features achieved 69% accuracy. Chapter Verbally interpret and research the symptoms of COVID-19 disease; very close to being accepted by the speakers. Each sentence of the mel filterbank properties for each phoneme is represented as a support vector. Extract features of COVID-19 speech from normal 12 speech using a binary classifier. A small video file collected from YouTube shows 88.6% accuracy and 92.0% accuracy.On this data, the SVM classifier can achieve an F1 Score of 7%. Further analysis showed that the two groups could be better distinguished by certain phonemes (pauses, midtones, and nasal sounds).

Radiologic changes in CT images of Chinese COVID-19 patients were determined. In this study, he used deep learning to extract graphic images of COVID-19 from CT scan images he developed as an alternative diagnosis method. They also collected scanned images of patients with confirmed COVID-19, as well as patients diagnosed with pneumonia. The results of their work provide proof of principle for the accuracy of COVID-19 predictions using AI . This study uses CT scan images, unlike our study, as we use clinical features and laboratory results for prediction.

This study describes epidemiologic & demographical & clinical & laboratory & radiological & clinical data from Zhongnan Hospital in Wuhan, China. Data were analysed and recorded for disease follow-up . The authors provide a better understanding of the radiological and clinical data that can be used to predict COVID-19 in our model.

A new framework has been proposed for testing the coronavirus using smartphones. Artificial intelligence systems are designed to collect data from multiple sensors to predict the level of lung inflammation as well as the spread of disease . The proposed methodology provides CT scan images as an important method for predicting COVID-19. The framework is based on multiple readings from multiple sensors regarding COVID-19 symptoms.

If not diagnosed and treated early, Covid-19 can cause many problems. These problems can be avoided. Some of these problems include nerve and heart disease, kidney damage and disease, slow wound healing, hearing loss, skin problems, and immune system problems. Early detection of Covid-19 is very important for maintaining a healthy life. In order to look for a answer to this important question, only recently developed education systems have been put into practice. 15 In the above article, various estimates have been developed using CT scan images and symptoms to predict COVID-19, risk of death and transmission in different countries. Based on the available information, there is not much evidence to predict the use of clinical data.

The project will use machine learning techniques to predict COVID-19 based on clinical data from COVID-19 patients. It also determines which features affect the model's prediction.

III. PROBLEM STATEMENT

It is essential to have an accurate diagnosis of COVID-19 in order to diagnose it early, treat it effectively and have better patient outcomes. However, the development of a reliable prediction model for Covid-19 is difficult due to the complexity and heterogeneity of the disease, as well as the number of variables that contribute to the genesis of the disease.

Traditional machine learning algorithms could struggle to deal with high-dimensional and nonlinear patterns and miss out on capturing the subtle relationships that exist in health da ta. Consequently, there is a need for a reliable and accurate prediction model that is able to effectively utilize the information available in health databases. Additionally some machine learning algorithm such as multilayer perceptron

(MLP), extreme machine learning (EML), and support vector classification (SVC) have shown promising results in various fields; however, each approach has certain limitations that it must adhere to. However, there is a possibility that MLP suffers from computational inefficiencies despite its power in capturing complicated nonlinear interactions.

EML provides simpler learning frameworks for better efficiency, but may lack essential pattern linking in the process. On the other hand, SVC is able to handle high-dimensional data and non-linear decision limits, but may not fully represent complicated connections. Therefore, in order to achieve a high level of accuracy in the prediction of Covid-19, it is necessary to design a hybrid machine learning model that incorporates the advantages offered by each of these techniques. Additionally, the availability of a complete health database that includes relevant clinical and demographic variables provides value to the process of developing a reliable predictive model. The use of health database information, on the other hand, presents additional obstacles such as dealing with missing values, standardizing characteristics, and dealing with class imbalance issues. In order to guarantee the accuracy of the prediction model 25 as well as its applicability to the wider population, it is necessary to pre-process and integrate the health data in an efficient way.

Therefore, the challenge is to design a hybrid machine learning algorithm that correctly predicts Covid-19 using a comprehensive health database and successfully combines MLP, EML and SVC. This is a problem that needs to be solved. The purpose of this project is to improve the accuracy of the prediction of Covid -19 while overcoming the limits given by specific algorithms. In addition, the constructed model should be able to handle difficulties that arise in data preparation, such as missing values and class imbalance, to guarantee accurate predictions in practical healthcare applications. Ultimately, the vision is to improve patient's results and quality of life by addressing these barriers, with the intention of providing clinicians with a reliable and accurate tool for early identification and successful treatment of Covid-19.

In conclusion, the prediction of COVID-19 disease using machine learning methods has the potential to transform healthcare by enabling early identification, individualized therapies, and successful treatment of the disease.

This would be possible due to increased forecasting accuracy. It is possible to build accurate and reliable prediction models for the healthcare industry by evaluating large-scale health information and harnessing the power of MLP, NAÏVE BAYES and SVC algorithms. These models

will contribute to better patient care and improved health outcomes.

IV. METHODOLOGY

Implementation: The following are the processes that make up the approach to building a hybrid machine learning model for predicting Covid-19 using the Pima Health data set and including Principal Component Analysis (PCA). This model will be used to predict whether or not a person has covid. The Pima health dataset, which includes relevant clinical and demographic information, is retrieved and checked for missing values or outliers. This is the initial stair in the data preprocessing process. Imputation methods such as mean imputation and regression imputation can be used to fill in the blanks after missing data. Outliers can be dealt with by removing them or replacing them with more acceptable values. In addition, the data set is divided into a character matrix (X) and a target variable (y), where the presence or absence of Covid-19 represents the target variable. PCA, or principal component analysis & is done to minimize the number of dimensions contained in the element matrix. PCA is a method for reducing the dimensionality of data by transforming the initial characteristics into a new set of variables called principle components that are uncorrelated. These components represent the largest amount of intrinsic variability in the data. The dimensionality of the feature matrix can be reduced while preserving the information most important to understanding the problem, if only those primary components that can account for a significant portion of the variation are retained. In order to build a hybrid machine learning model, the data must first be preprocessed and then subjected to principal component analysis. The MLP, EML, and SVC algorithms were combined in this model to make the most of each of their respective capabilities. The MLP technique, which is based on neural networks, is able to capture the complicated nonlinear correlations present in the data. EML, on the other hand, provides a learning framework that is easier to understand and increases the efficiency of 30 computational operations. A sophisticated classification method known as SVC is capable of handling highdimensional data as well as non-linear decision boundaries. Training and evaluation: The pre-processed and reduced feature matrix is divided into training and test sets in addition to the target variable. Then, MLP, EML, and SVC training algorithms are applied to the training set to train the hybrid model. To achieve the best performance, the model parameters are fine-tuned using methods such as grid search and random search. The trained hybrid model is then used to generate predictions on the test set, after which evaluation metrics such as accuracy and precision and recall and F1 score are calculated. This allows the model to be evaluated. In order

to find out whether the hybrid model is better than the individual MLP, EML, and SVC models, the performance of the hybrid model is evaluated and compared in addition to other baseline models. 21 Analysis and Interpretation of Findings Here, the results of the hybrid model are studied and explained, including measures of accuracy and performance. A feature importance analysis can be performed to understand the factors that have the most significant impact on forecasts. This study contributes to the process of explaining how the model reached its conclusions and offers new insights into the components that play a role in predicting Covid-19. It is essential that the model is interpretable by medical professionals so that they can understand and trust the predictions made by the model when applied to real situations in the medical field. Overall, this methodology including data preprocessing, principal component analysis (PCA), hybrid model development, training, and evaluation and interpretation of results provides a comprehensive approach to accurately predict Covid-19 using the Pima dataset while incorporating the strengths of the MLP, EML, and SVC algorithms.

To ensure that the data is ready for modelling, data pretreatment is performed. Scaling the features and standardizing them to have mean zero and variance one is done using scikit-learn' s Standard Scaler. This phase is necessary for algorithms that are sensitive to the size of the features they are analyzing.

Multilayer Perceptron (MLP) is a special type of artificial neural network (ANN). It consists of many layers of connected nodes, which are also called neurons. Powerful Common techniques are used for supervised learning problems such as classification and regression top of all. In multilayer perceptron (MLP), each neuron in each layer receives one input It receives signals, applies a nonlinear activation function to those signals, and sends the modified data. It is applied to nerve cells in its upper layers. 24 This allows the model to capture complex situations. Nonlinear correlation between input characteristics and input variables has been targeted. This This algorithm changes the neuron weights and biases to reduce the number of neurons. The error between the expected output and the actual output. It is known that Due to its ability to learn and estimate complex functions, Do things with non-linear interactions, like predicting diabetes.

Extreme machine learning, also known as ELM, is a machine learning paradigm that combines: Advantages of artificial neural networks and extreme random projection methods Machine learning is also called ELM. are ELMs Simplified for efficient training and prediction. No ELM required Unlike building iterative optimization techniques

such as post-release work, A typical neural network, instead, starts with weights that link the input layer to layer. Changed functions are created a hidden layer that uses a non-linear activation function to apply to the weights inputs. Then, linear regression or classification is performed using the output layer creation use of changed functions. The basic concept of ELM is the idea that effective capture of non-linear connections can be achieved by randomly projecting the input data into high-dimensional environments. ELM is suitable for use with large datasets because to its high level of computational efficiency, as well as its fast training and prediction Time frame. In the field of Covid-19 prediction, ELM can provide a useful method efficiently capture complex correlations in health data while minimizing their quantity and the complexity for the calculations that are required for this.

Support vector classification is a widely used approach which can be used to classify tasks involving binary or multi-class categories. It is based on the idea behind finding an ideal hyperplane that divides the data points different classes with the largest margin. Using the kernel method, SVC can do this successfully handle high-dimensional data as well as decision boundaries that are non-linear. The kernel function takes the input characteristics and transforms them into a space with a higher dimension. This makes it possible to distinguish between data points that could not be differentiate linearly using the initial character space. SVC performance can be improved by fine-tuning it hyperparameters such as kernel function selection and regularization parameter, Besides. In the field of diabetes prediction, SVC is able to capture complicated linkages and handle non-linear decision limits, which positions it as an a suitable method for proper categorization of persons at risk of diabetes.

Naive Bayes is one of the very simple and powerful algorithms for classification based on Bayes theorem with the assumption of independence between predictors. The Naive Bayes classifier assumes that the presence of an element in a class is unrelated to any other element. Naive Bayes is a classification algorithm for binary and multi-class classification problems. Based on prior knowledge of the conditions that may be associated with an event, Bay es' theorem describes the probability of an event the conditional probability can be found in this way Suppose we have a hypothesis (H) and evidence (E).Suppose we have a dataset of weather conditions and a corresponding target variable "Play". So using this data set we have to decide whether to play or not on a particular day based on the weather conditions. To solve this problem we need to do the following steps: 25

• Transform the mentioned dataset into frequency tables.

- Generate a probability table by finding the probabilities of the given properties.
- Now use Naïve Bayes theorem to calculate the probability.

V. CONCLUSION & FUTURE WORK

The healthcare industry collects a huge amount of data that contains some hidden information, which is useful for effective decision making. Some advanced data mining techniques are used to provide suitable results and take effective decisions on the data. In this study, an effective Covid-19 disease prediction system (ECDPS) using neural networks is developed to predict the risk level of heart diseases. The developed model uses 12 medical parameters such as age, gender, blood pressure, cholesterol and obesity to predict. ECDPS predicts the likelihood that patients will develop heart disease. It enables important knowledge to be established, such as relationships between medical factors related to heart disease and patterns. We used a multilayer perceptron neural network with backpropagation as a training algorithm. The obtained results showed that the proposed diagnostic system can effectively predict the level of risk of heart diseases. In this research, a systematic literature review was conducted to identify a suitable algorithm for predicting COVID-19 in patients. No clean evidence has been found that summarizes a single algorithm as the appropriate technique for prediction. Therefore, a set of algorithms was chosen which includes support vector machine (SVM) & artificial neural networks (ANN) & random forests (RF). The selected algorithms were trained with the patient's clinical information. To evaluate the accuracy of the machine learning models, each algorithm is trained with datasets of different numbers of patients. The trained algorithms were evaluated using the performance accuracy metric. After analysing the results, Random Forest (RF) showed better prediction accuracy compared to Support Vector Machine (SVM) and Artificial Neural Networks (ANN). The trained algorithms were also evaluated to find features that influence the prediction of COVID-19 in patients. To gain the trust of health industry professionals, it is important to improve the interpretability and explain the predictive power of the model. An important goal is to develop methods that provide a clear explanation of the decisions that the model makes, emphasizing the underlying characteristics and patterns that are responsible for the forecasts. There is a lot of room for machine learning in healthcare. For future work, it is adviced to work on calibrated and ensemble methods that could solve unpredictable problems faster with better results than existing algorithms. Using various sensors and features, an AI-based application can also be developed to identify and help diagnose diseases. 30 Since health care prediction is a crucial area of the future, a prediction system can be developed that could detect the possibility of new disease outbreaks that could harm humanity by taking socio-economic and cultural factors into account. Exploring the feasibility of integrating the hybrid model into clinical practice Implementation into clinical practice Further exploration of the feasibility of integrating the hybrid model into clinical practice and health care systems is needed. Collaborate with healthcare professionals to evaluate the impact of the model on patient decision-making and outcomes and resource allocation. In doing so, be sure to consider variables such as ethical concerns, regulatory constraints, and implementation issues. The hybrid machine learning model can be continuously improved by addressing these future research topics, leading to increased accuracy of diabetes predictions and contributing to advances in personalized healthcare and disease management. Investigating more complex methods to improve the parameters and hyperparameters of the hybrid model is an important step in the algorithm optimization process. For this, it may be necessary to use more complex optimization methods such as Bayesian optimization or evolutionary algorithms to find the ideal combination of parameters that would increase the performance of the model to its full potential. Ensemble Learning Techniques: Research and development of innovative ensemble learning techniques that successfully incorporate MLP, EML and SVC predictions should be the primary focus of this project. To achieve this, several methods such as stacking, boosting or bagging may need to be explored to further improve the accuracy and robustness of the model. Incorporating Domain Information The model development process should include domainspecific information as well as domain expert knowledge. Expert input can be beneficial in many ways, including identifying significant characteristics, improving preprocessing methods, and providing better knowledge of the underlying processes that contribute to the prediction of diabetes. The implementation results show that the MLP method combined with PCA and preprocessing approaches achieves maximum accuracy, which is promising for reliable diabetes prediction. The results were obtained from the implementation. In addition, several machine learning algorithms are used, including K-Nearest Neighbor (KNN), Naive Bayes, and Logistic Regression, which provide insight into different categorization strategies.

REFERENCES

[1] Coronavirus Disease (COVID-19) - events as they happen. Library Catalog: www.who.int.
[2] Countries where Coronavirus has spread World meter.www.worldometers.info.
[3] COVID-19 situation reports. Library Catalog: www.who.int.

[4] Diagnosis of covid-19 and its clinical spectrum dataset. url=https://kaggle.com/einsteindata4u/covid19.

[5] WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020. Library Catalog: www.who.int.

[6] WHO EMRO | Questions and answers | COVID-19 | Health topics.

[7] Support Vector Machine learning algorithm with example and code, January 2019. Library Catalog: www.codershood.info Section: Machine learning.