# Survey On Novel Implementation of Covid-19 Classification Model

## Mrs. Thilagavathi N[1], Kaushika M[2], Keerthana M[3], Chandralekha T[4]

[1]Assistant Professor, Department of Information Technology, Sri Manakula Vinayagar Enginerring College, Puducherry-605107

[2,3,4]UG Students, Department of Information Technology, Sri Manakula Vinayagar Engineering College, Puducherry-605107

-----------------------------------------------------------------***---------------------------------------------------------------------

**Abstract -** *Technology advancements have a rapid effect on every field of life, be it medical field or any other field. Artificial intelligence has shown the promising results in health care through its decision making by analyzing the data. COVID-19 has affected more than 100 countries in a matter of no time. People all over the world are vulnerable to its consequences in future. It is imperative to develop a control system that will detect the coronavirus. One of the solutions to control the current havoc can be the diagnosis of disease with the help of various AI tools. In this project, we classified textual clinical reports into four classes by using classical and ensemble machine learning algorithms. Feature engineering was performed using techniques like Term frequency/inverse document frequency, Bag of words and report length. These features were supplied to traditional and ensemble machine learning classifier like Yolo classifier and showed better results than other ML algorithms by having more than 95 testing accuracy. In future recurrent neural network can be used for better accuracy. The recent Coronavirus Disease 2019 (COVID-19) pandemic has placed severe stress on healthcare systems worldwide, which is amplified by the critical shortage of COVID-19 tests. Attributes such as age, blood pressure, thickness of the artery, Respiratory Rate etc. are fed into the Deep Learning means algorithm which is used to predict risk of Coivd-19 in a person.*

*Keywords: yolo, fizzy, hospitality, ML.*

## INTRODUCTION:

The novel Coronavirus designated SARS-CoV-2 appeared in December 2019 to initiate a pandemic of respiratory illness known as COVID-19 which proved itself as a tricky illness that can emerge in various forms and levels of severity ranging from mild to severe with the risk of organ failure and death. From mild, self-limiting respiratory tract illness to severe progressive pneumonia, multiorgan failure, and death had described by D. Wang et al (may 2019) . With the progress of the pandemic and rising number of the confirmed cases and patients who experience severe respiratory failure and cardiovascular complications, there are solid reasons to be tremendously concerned about the consequences of this viral infection . T. Guo et al (may 2020) has Determined appropriate approaches to reach solutions for the COVID-19 related problems have received a great deal of attention. However, another huge problem that researchers and decision-makers have to deal with is the ever-increasing volume of the date, known as big data, that

challenges them in the process of fighting against the virus. This justifies how and to what extent Artificial Intelligence (AI) could be crucial in developing and upgrading health care systems on a global scale. P. Hamet and J. Tremblay (April 2017) stated AI has been recently attracted increasing research efforts towards solving the complex issues in a number of fields, including engineering. M. Jamshidi, A. Lalbakhsh, S. Lotfi,(March 2020) probed critical situation like this necessitates mobilization and saving medical, logistic and human resources and AI can not only facilitate that but can save time in a period when even one hour of the time save could end in saving lives in all locations where Coronavirus is claiming lives. M. B. Jamshidi, N. Alibeigi (November 2018) described With the recent popularity of AI application in clinical contexts, it can play an important role in reducing the number of undesired deletions as well as improving the productivity and efficiency in studies where large samples are involved S. Huang, J. Yang(Aug 2019),stated the higher degrees of accuracy in prediction and diagnosis are intended P. Rashidi and A. Bihorac(Feb 2020) has Utilized big data can also facilitate viral activity modelling studies in any country. The analyses of results enable health care policymakers to prepare their country against the outbreak of the disease and make well-informed decisions D. S. W. Ting, L. Carin (May 2020) stated Nevertheless, while treatment strategies, crisis management, optimization and improvement diagnosis methods, such as medical imaging and image processing techniques could take benefit from AI which is potentially capable of helping medical methods, it has not been desirably employed and well-appropriated to serve health-care systems in their fights against COVID-19. For instance, one area that can take special advantage of AI's useful input is image-based medical diagnosis through which fast and accurate diagnosis of COVID-19 can take place and save lives W. Naudé (April 2020) stating AI techniques to deal with COVID-19 related issues can fill the void between AI-based methods and medical approaches and treatments. S. Luminari(Oct 2019) describes AI specialists' use of AI platforms can help in making connections between various parameters and speed up the processes to obtain optimum results. In this paper, our team relies on the findings of the most recent research focusing on COVID-19 and its various challenges to generalize and suggest a variety of strategies relevant but non-limited to high-risk groups, epidemiology, radiology and etc.

## 2. METHODOLOGIES:

### i. Data analysis:

Data Analysis is a way of thinking about information from social events and then set it up for major conferences. Information analysts discuss the use of notable methods related to the description and control of information. Every one of these bits of knowledge permits the organizations to define better procedures and to settle on remotely enhanced choices. Data Analysis is characterized as a procedure of cleaning, changing, and displaying data to find valuable data for business basic leadership. The motivation behind Data Analysis is to extricate valuable data from information and getting the preference reliant on the data analysis. Similarly, Data Analysis is a procedure of examining, purging, changing and dis- playing data with the objective of finding helpful data and establishing basic leadership. Data analysis has many aspects and methods. They combine different methods under many different names and used in different fields of business, science, and sociology. In today's business world, data analysis is responsible for making progressive logical choices and assistance to organizations work more successfully. Data analysis is a procedure of gathering and demonstrating information with the objective of finding the necessary data. The outcomes are conveyed, proposing ends, and supporting basic leadership. Occasionally, data perception depicts the simplicity of finding valuable examples in the data for the data. Data analysis along with Data modelling statements indicate the consequent.

This prompts progressively insightful business moves, higher advantages, profitable maneuvers, and energetic customers. The idea is to share the business prospects in a better way soon and then use it with an examination idea. Information expands at a fast speed and the pace of improvement of information is high.
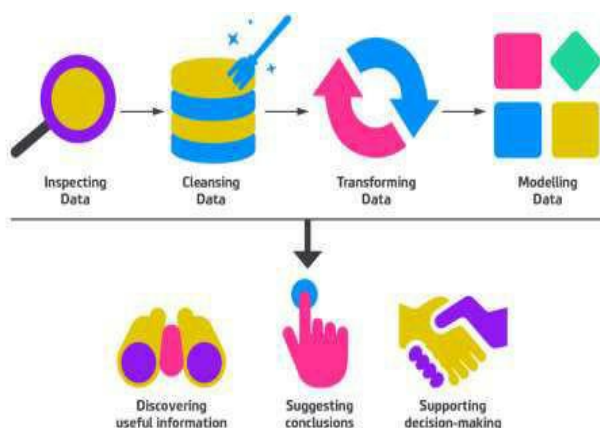


Fig.1      Data Analysis

Information analysis is a practice for constructing original data and solving various models in the data through numerical estimation and calculation. Analyzing data facilitates to create new information and accessible various

bits of information. Obtaining data from an organization's database or deleting data from external sources for research is one of the major occupations of any data analyst. Clearing information is the basic stage of the entire data classification process, and a stupid and unimproved way to review, identify, and use the data. When investigating deterministic data to select key decisions, data analysts should start with prudent data cleansing methods. It is important to check the phenomena established by data cleansing, and cleansing incorporates data deletion operations that may destroy score or organize information in a single way. Apparently, this activity is necessary for any data analyst. Data testing is a profession that studies vague real data elements to meet specific requirements. This is a way to evaluate data using logical reasoning and real predicting to quickly view each data provided. People use quantitative tools to verify and dismantle data. Extensive data validation is dedicated to discovering examples, links, and models in confusing data sets. The information analyst will consider both the current time and the long- term situation. Standard examination affects to perceive how business has performed and anticipate where business maneuvers and practices will be held. Sufficiently, it will give reflections in respect of how it might change issues to move the business preference. Announcing makes a translation of crude data into information while revealing urges associations to screen their online business and became conscious of when data falls outside of predicted degrees. Remarkable reporting ought to raise issues about the business from its end customers.

### ii. Data science:

Data Science is a collective term for a set of progressively powerful fields that focus on extracting large amounts of information indexes and working to find new creative pieces of knowledge, models, technologies, and procedures. Data analysis depends on the order in which important experiences are added to help quickly expand the company's specialists. It is part of a broader strategy that be component of Information Science. It predicts the future based on past patterns.
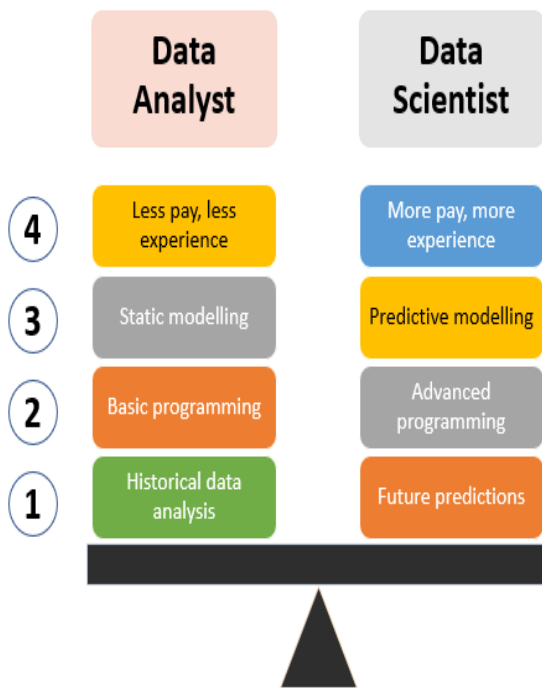
Fig.2 Comparison of data science with data analysis

Compared to Data Analysis and Data Science, the variety is displayed at this moment, and blows the problem on small and large scales. As mentioned above, fundamentally, Data Science is a broad inter- disciplinary field that covers the broader field of Data Analysis, and it is closely linked with the great processing of organized and unstructured information. On the other hand, Information Analysis is a miniaturized field that penetrates the explicit components of business activities, and its ultimate purpose is to inform the departmental model and rationalize the form. Focusing on organized information, there are many information analyses models that show the actual situation and its impact on the company. Although these two instructions analyze a wide range of companies, specializations, ideas, and practices, Data Science is often used in important areas such as business exams, web index design, and self-sufficiency, such as knowledge ability AI and AI ML. Data Analysis is an evolving and progressive idea; though, this area of competence or computerized data innovation is often used within human services, retail, gaming and travel companies to quickly respond to company difficulties and goals.

## 3. COVID-19:

N. Chen[June 2020] described COVID-19 is the disease caused by a new coronavirus called SARS-CoV-2. WHO first learned of this new virus on 31 December 2019, following a report of a cluster of cases of 'viral pneumonia' in Wuhan, People's Republic of China. Among those who develop symptoms, most (about 80%) recover from the disease without needing hospital treatment. About 15% become seriously ill and require oxygen and 5% become critically ill and need intensive care. Complications leading to death may include respiratory failure, acute respiratory distress syndrome (ARDS), sepsis and septic shock, thromboembolism, and/or multiorgan failure, including injury of the heart, liver or kidneys. In rare situations, children can develop a severe inflammatory syndrome a few weeks after infection. *Some people who have had COVID-19, whether they have needed hospitalization or not, continue to experience symptoms, including fatigue, respiratory and neurological symptoms. WHO is working with our Global Technical Network for Clinical Management of COVID-19, researchers and patient groups around the world to design and carry out studies of patients beyond the initial acute course of illness to understand the proportion of patients who have long term effects, how long they persist, and why they occurs. These studies will be used to develop further guidance for patient care.* The time from exposure to COVID-19 to the moment when symptoms begin is, on average, 5-6 days and can range from 1-14 days. This is why people who have been exposed to the virus are advised to remain at home and stay away from others, for 14 days, in order to prevent the spread of the virus, especially where testing is not easily available. In most situations, a molecular test is used to detect SARS-CoV-2 and confirm infection. Polymerase chain reaction (PCR) is the most commonly used molecular test. Samples are collected from the nose and/or throat with a swab. Molecular tests detect virus in the sample by amplifying viral genetic material to detectable levels. For this reason, a molecular test is used to confirm an active infection, usually within a few days of exposure and around the time that symptoms may begin.

## 4. MACHINE LEARNING:

K. Liu, Y.-Y. Fang [April 2019] detailed the Machine learning is the ability of machine, computers to learn itself and acts like human beings and it improves it's learning capability in every input by feeding information in the form of observation with the real world interaction. In Machine Learning a computer program is assigned to perform some tasks and it is said that the machine has learnt from its experience if its measurable performance in these tasks improves as it gains more and more experience in executing these tasks. So the machine takes decisions and does predictions and forecasting based on data. Machine Learning is applied in wide variety of fields namely: robotics, virtual personal assistants (like Google), computer games, pattern recognition, natural language processing, data mining, traffic prediction, online transportation network. It is a subfield of Computer Science. Resurging interest in machine learning is due to the same factors that have made data mining and Bayesian analysis more popular than ever. Things like growing volumes and varieties of available data,

computational processing that is cheaper and more powerful, and affordable data storage. All of these things mean it's possible to quickly and automatically produce models that can analyze bigger, more complex data and deliver faster, more accurate results even on a very large scale.
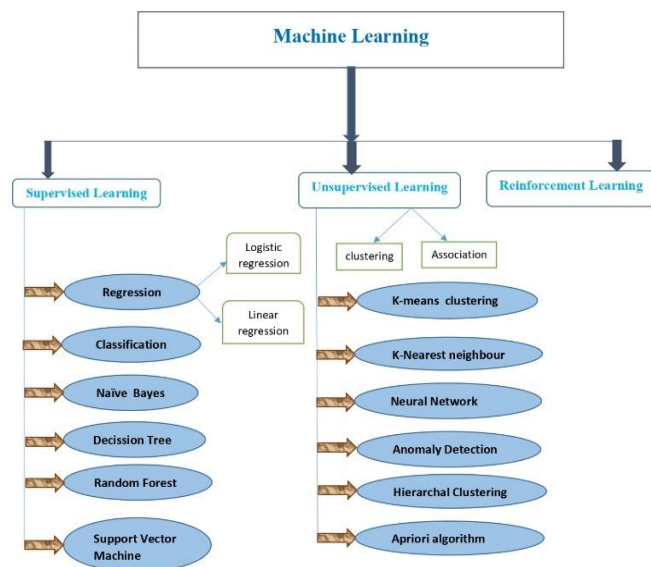


Fig 3: Machine learning classification diagram

Machine learning has 3 types, they are:

## a. Supervised learning:

In supervised learning, it uses the labelled datasets to train algorithm to classify data and it also predicts the accuracy of outcomes. This technique is accomplished by providing training, input and output patterns to the systems. Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. One good example for supervised learning is text classification. Supervised ML techniques utilizes from what it is has gained knowledge from the previous and present data with the help of labels to forecast events. This approach initiates from the training process of dataset, ML develop an inferred function to foresee the output values. The system is capable of providing results to an input data with adequate training process.

Supervised learning has 6 types of algorithms.

## i. Regression:

It is a supervised technique where regression models are used to analyze and predict the continuous value, like admission of the patient, is one of the common example. Regression is an approach of supervised learning. In Regression we have the labeled datasets and the output variable value is determined by input variable values - so it is the supervised learning approach. The most simple form of regression is linear regression where the attempt is made to fit a straight line (straight hyperplane) to the dataset and it is possible when the relationship between the variables of dataset is linear. This Regression Analysis technique involving multiple variables can be used in property valuation, car evaluation, forecasting electricity demand, quality control, process optimization, and medical diagnosis.

It has 2 types, they are:

## Logistic Regression:

Logistic regression is the technique borrowed by machine learning from statistic. Logistic regression is used to deal a classification problem. It is a supervised learning algorithm that includes more dependent variables. The response of this algorithm is in the binary form. Logistics regression can provide the continuous outcome of a specific data. This algorithm consists of statistical model with binary variables. It gives the binomial outcome as it gives the probability if an 37 event will occur or not (in terms of 0 and 1) based on values of input variables. For example, predicting if a tumor is malignant or benign or an e-mail is classified as spam or not are the instances which can be considered as binomial outcome of Logistic Regression. There can be multinomial outcome of Logistic Regression as well.

## Linear Regression:

Linear regression is the basic form of regression analysis, it assume that there is linear relationship between predictor. Linear regression has the advantage that it is easy to understand and it is also easy to avoid over fitting by regularization. Also we can use SGD to update linear models with new data. Linear Regression is a good fit if it is known that the relationship between covariates and response variable is linear. It shifts focus from statistical modeling to data analysis and preprocessing. Linear Regression is good for learning about the data analysis process. However, it is not a recommended method for most practical applications because it oversimplifies real world problems.

## ii. Classification:

It is used to classify the set of given data. The data can be structured format or unstructured format. It gives accurate prediction in which set the data is been given. Classification is a process of data partition according to the features of a

group of data objects, and has been widely studied in statistics, machine learning, neural network and expert system. Now it has become one of the key research aspects of data mining. In fact, the classification is a two-step process. In the first step, a model is built to describe a predetermined set of data classes or concepts. In the second step, the predictive accuracy of the model is estimated, if the accuracy of the model is considered acceptable, the model can be used to classify future data objects.

### iii. Naive Bayes:

Naive bayes is one of the simple and most effective algorithm which help in building the fast machine learning models that can make quick prediction. Naive Bayes is the most up-front and rapid classification algorithm, which is suitable for an enormous block of data. There are varied applications such as sentiment analysis, text categorization, spam filtering and recommender systems, where NB classifier is being used. It deals with the datasets which have the highest dimensionality Bayes theorem of probability is used for predicting the unknown classes. Naive Bayes is straightforward and easy to implement algorithm. Because of which, when the quantity of data is sparse it might out perform more complex models

### iv. Decision Tree:

A decision tree is a flow chart like structure in which each internal node represents a test on a feature, each leaf node represents a class label and branches represents conjucations on features. We can represent any boolean function on discrete using the decision tree. Decision trees are constructed as an algorithmic approach that identifies ways to split a data set based on different conditions. It is one of the most widely used method for supervised learning. The aim is to build a prototype that predicts the worth of a target variable by learning straightforward decision tree instructions and it does not require any parameter setting, and therefore it is appropriate for discovery of the knowledge. Decision trees performs classification without requiring much computation. Decision trees is capable to handle continuous as well as categorical variables.

### v. Random Forest:

Random forest is a supervised learning algorithm. Like the forest it builds, is an ensemble of decision trees, and it is usually trained with the "bagging" method. The general idea of bagging method helps to increase the overall result by the combination of learning models. which is used for both classification and regression also. However, it is mainly used for classification purposes. The name itself is suggested that it is a forest, a forest is a group of trees similarly in a random forest algorithm we will have trees these trees are the decision trees. If we have a higher number of decision trees prediction results will be more accurate. Random forest algorithm works this way at; first it will collect random samples from the dataset and then it will create decision

trees for each sample from those available trees we will select the tree which will produce the best prediction results.

### vi. Support Vector Machine:

Algorithm which can be used for both classification and regression challenges is said to be support vector machine learning. In this type of algorithm each data is plotted depending upon the number of features you have with the value of particular co-ordinate. A support vector machine is a supervised learning technique in machine learning algorithms. If you give any labeled training data to support vector machine algorithms , it will produce a classifier that will divide the labeled data into different classes. In the one-dimensional (1D) space, this classifier is called a point. In two-dimensional (2D) space, this classifier is called a line. In three-dimensional (3D) space, this classifier is called a plane. In four-dimensional(4D) or more space, this classifier is called a hyperplane.

### b. Unsupervised Learning:

unsupervised learning is a self-learning technique, it uses machine learning algorithms to discover the features of the input by its own and no prior set of categories is used. Unsupervised ML techniques are employed when the training data is non-classified and not labeled. It analyses how the system can deduce a function to explain the hidden patterns from the unlabeled data. The system does not identify the proper output, but it discovers the data and writes observations from dataset to find hidden patterns from unlabeled data.

 It has 2 types, they are

### i. Clustering:

Clustering is process of grouping similar entities together. This technique is used to find and group similar data points together. It is contiguous regions of more than one-dimensional space comprising comparative points of high density. In image breakdown, clustering is the order of arrangement of pixels conferring to more or less features like intensity. Under hard clustering, data elements fit into one cluster simply and the membership value of belongingness to a cluster is precisely one. Under soft clustering, elements of data fit into more prominent than the single cluster and the membership value of belongingness to the cluster varies from 0 to 1.
It has 6 types of algorithms.

### K-Means Clustering:

In K-mean Clustering, it groups the data which starts with k as input. It helps to find how many clusters you want by placing K centroid in random location. This Algorithm is frequently used for solving clustering problem. It is a form of unsupervised learning. It has the following advantages: it is computationally more efficient than hierarchical clustering

when variables are huge. With globular cluster and small k it produces tighter clusters than hierarchical clustering. Ease in implementation and interpretation of the clustering results are the attraction of this algorithm.

## K-Nearest Neighbour:

It is one of the simplest machine learning technique. It classifies the data based on the similarities between the new data and available data and put the case into most similar available category. K Nearest Neighbor (KNN) Algorithm is a classification algorithm. It uses a database which is having data points grouped into several classes and the algorithm tries to classify the sample data point given to it as a classification problem. KNN does not assume any underlying data odistribution and so it is called non-parametric. Advantages of KNN algorithm are the following : it is simple technique that is easily implemented. Building the model is cheap. It is extremely flexible classification scheme and well suited for Multi-modal classes. Records are with multiple class labels.

## Neural Network:

Neural networks are subset of machine learning. Here datasets are modelled in complex pattern using hidden layers and non-activation layers of algorithm which computer can understand. The supervised learning is used by neural network which classifies the input information into the desired product it is mainly designed for computational purposes; the main theme of this model is to do a job faster than the traditional model. This model is similar to the biological structure of neurons in the human brain. How the neurons connect in the brain the same way here, also neurons (nodes) will connect. This model consists of a vast number of interconnected elements (neurons) working collectively to perform a task. A single layer neural network is called a perceptron it gives us a single output.

## Anomaly Detection:

In unsupervised machine learning requires training data because it assumes two things i.e, small part of the data as anomaly and any anomaly is different from the normal samples. Based on these assumption similar data are clustered together and data points which are far off from the points are considered as anomalies. Anomaly detection (AD) is designed to find targets that are spectrally distinct from their surrounding neighborhood. Since AD must be performed in a completely blind environment, two issues must be addressed, one is, how to find anomalies without using any prior knowledge and another one how to distinguish one anomaly from another once anomalies are detected.

## Hierarchical Detection:

In Hierarchical clustering it considers all data points as their own cluster. It combines the two nearest data and merges as one point. Hierarchal Clustering is too expensive. It is for this reason that hierarchal clustering has been avoided in some application. This method that built clusters top-down that is they begin with all item in one large cluster and break down that cluster over a number of iterations. Hierarchal clustering to be useful in many situations but having a cost of O(n3 ). This makes it too expensive to use in many real world applications. Hierarchal Clustering also Known as Hierarchal cluster analysis. The advantage of Hierarchal clustering is no pre-set number of clusters.

## Apirori Algorithm:

This algorithm is designed to work on the databases that contain transactions and this is done with the help of association rule where it develops frequent item sets of data. It also ensures that how strongly or weakly two objects are connected. Apriori algorithm is a highly popular pattern-finding algorithm in data mining technique. This algorithm is aimed at finding an itemset combination which has a certain value of frequency according to the desired criteria or filter. This algorithm is proposed by R. Agrawal and R. Srikant. The result of apriori algorithm can be used to help the management make decisions. Apriori algorithm do an iterative approach known as level-wise search. apriori algorithm can increase the efficiency of data calculation analysis. Additionally, another study mentions that apriori based algorithm has better speed in managing greater data.

## ii. Association:

Association learning is a Rule based machine learning. Un this technique, it checks the dependency of one data with another data by measuring degree of similarity, hidden correlation in databases. It also provide interesting association rules for new searches. An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

## c. Reinforcement Learning:

Reinforcement learning is one of the area in Machine learning. It is mainly about maximizing reward for a suitable action at a particular situation. To find the best possible behavior or path or results it is employed by various software and machines. Reinforcement Learning presupposes that agents can visit a finite number of states, and when visiting a state, a numerical reward will be collected; negative numbers represent punishments. Each state has a changeable value attached to it. From every state, subsequent states can be reached by actions. The value of a given state is defined by the averaged future reward that can be accumulated by selecting actions from this particular state. Reinforcement learning having two learning model Q-learning and Markov decision process.

## 5. LITERATURE SURVEY:

### a. Supervised Learning:

### i. Regression:

Furqn rustam et al (25 May 2020), described using Ml models to forecasting of COVID-19 spread focusing on the number of new positive cases, the number of deaths, and the number of recoveries.The forecasting has been done by using four Supervised Machine Learning Models approaches that is linear regression, Least absolute shrinkage and selection operator (LASSO), support vector machine (SVM), and exponential smoothing. Linear regression is used to find out the relationship between independent and dependent variables and also for forecasting. LASSO is used to reduce the number of errors in this model. SVM solves the regression problems using a linear function. ES is a very simple powerful time series forecasting method specifically for univariate data. The Issue is SVM produces poor results in all scenarios because of the ups and downs in the dataset values. It was very difficult to put an accurate hyperplane between the given values of the dataset.

Pankaj et al. (22 January 2021). Stated Seven classifier algorithms have been applied in this research such as artificial neural network, C5.0, Chi-square Automatic interaction detector, logistic regression, linear support vector machine with penalty L1 & with penalty L2 and random tree. They have developed a model to predict CKD disease in patients. The performance of the model was tested on both all attributes and selected features. predict Chronic Kidney Disease based on full features and important features of Chronic Kidney Disease data. The drawback is logistic and KNN did not give suitable results.

### ii. Classification:

Jian Ping Li et al (09 June 2020) proposed an efficient and accurate system to diagnosis heart disease and the system is based on machine learning techniques. The system is developed based on classification algorithms includes Support vector machine, Logistic regression, Artificial neural network, K-nearest neighbor, Naïve bays, and Decision tree while standard features selection algorithms have been used such as Reliefs, Minimal redundancy maximal relevance, Least absolute shrinkage selection operator and Local learning for removing irrelevant and redundant features. All FS algorithms results show that the feature Fasting blood sugar (FBS) is not a suitable heart disease diagnosis.

### iii. Naïve Bayes:

L.J.Muhammad (05 November 2020).explained by using supervised machine learning techniques orgistic regression, decision tree, support vector machine, naive Bayes, and artifcial neutral network. In this work, supervised ML techniques are used to develop predictive models for the

COVID-19 infection, using an epidemiology labeled dataset for positive and negative COVID-19 cases in Mexico with supervised learning algorithms. It shows the percentage of COVID-19 positive patients correctly by the models and covid negative patients did not identified accurately.

Haroon Ahmed et al (14 August 2020), described machine learning and deep learning algorithms help us for diagnoses and predictions of such types of breast cancer. The machine learning algorithms are Arificial neural network, Naive bayes, logistic regression decision tree. The Naïve Bayes provides the highest accuracy while calculating the probabilities of noisy data that is used as an input. The main purpose of this research is to review different machine learning and data mining algorithms that helped people for the prediction of breast cancer. Our main focus is to find out the most accurate and suitable algorithm for breast cancer prediction.

### iv. Decision Tree:

Priyanka sonar et al (29 august 2019), the aim of this analysis is to develop a system which might predict the diabetic risk level of a patient with a better accuracy. Model development is based on categorization methods as Decision Tree, ANN, Naive Bayes and SVM algorithms. Decision tree is used to split a data set based on different conditions. Support vector Machine goal is to determine the maximum-margin hyperplane which provides the greatest parting between the classes. Bayes theorem of probability is used for predicting the unknown classes. Artificial Neural Network (ANN) is tested after the training process to acquire the reaction of the network which states whether the disease is classified magnificently or not. The drawback of the SVM algorithm is that to achieve the best classification results for any given problem and ANN is difficult with dealing with big data with complex model. Require huge processing time.

Rahul Katarya et al. (04 August 2020) predicted the heart disease at the early stage using supervised machine learning techniques artificial neural network (ANN), decision tree (DT), random forest (RF), support vector machine (naive), naïve Bayes) (NB) and k-nearest neighbour algorithm. The data sets will be directly given to supervised machine learning algorithms, and supervised machine learning algorithms will perform according to their nature, and those algorithms will give some outputs. There are some common attributes which are used to predict the heart diseases are: Heart rate, age gender, ECG result, cholesterol and blood sugar. In this model the random forest did not give perform effectively.

### v. Random Forest:

L.J.Muhammad (05 November 2020) stated using supervised machine learning techniques orgistic regression, decision tree, support vector machine, naive Bayes, and artifcial neutral network. In this work, supervised ML techniques are

used to develop predictive models for the COVID-19 infection, using an epidemiology labeled dataset for positive and negative COVID-19 cases in Mexico with supervised learning algorithms. It shows the percentage of COVID-19 positive patients correctly by the models and covid negative patients did not identified accurately.

Pinar Cihan et al. (17 November 2020), fuzzy rule basing system (FRGS) used to predict the number of Covid-19 daily cases. Random Forest(RF) combines linguistic and numerical information for use in FRBS. The WM method has been implemented to forecast the number of verified cases in the Covid-19 outbreak. In the used dataset, total number of intensive care patients, total number of intubated patients, number of daily tests, number of daily recovered patients, number of daily mortality were used as input variables, and the number of cases was estimated. Due to rapid increase in covid cases so sometimes collapsed to predict the forecasting the patients.

## vi. Support Vector Machine:

The forecasting has been done by using four Supervised Machine Learning Models approaches that is linear regression, Least absolute shrinkage and selection operator LASSO, support vector machine (SVM), and exponential smoothing. Linear regression is used to find out the relationship between independent and dependent variables and also for forecasting. Support Vector Machine is used to reduce the number of errors in this model. SVM solves the regression problems using a linear function. ES is a very simple powerful time series forecasting method specifically for univariate data. The Issue is SVM produces poor results in all scenarios because of the ups and downs in the dataset values. It was very difficult to put an accurate hyperplane between the given values of the dataset.

L.J.Muhammad (05 November 2020)., described using supervised machine learning techniques orgistic regression, decision tree, support vector machine, naive Bayes, and artifcial neutral network. In this work, supervised ML techniques are used to develop predictive models for the COVID-19 infection, using an epidemiology labeled dataset for positive and negative COVID-19 cases in Mexico with supervised learning algorithms. It shows the percentage of COVID-19 positive patients correctly by the models and covid negative patients did not identified accurately.

## b. UNSUPERVISED LEARNING:

## i. K-Nearest Neighbour:

Chen et al (30 December 2019) stated the Chronic Kidney Disease (CKD) data set was obtained from the University of California Irvine (UCI) machine learning repository, which has a large number of missing values. CKD data set is difficult to maintain and values are missing. Using K-nearest

neighbor imputation to fill in the missing values in the data set, which could be applied to the data set with the diagnostic categories are unknown. KNN imputation was used to selects the K complete samples with the shortest Euclidean distance for each sample with missing values. However, in the process of establishing the model, due to the limitations of the conditions, the available data samples are relatively small, including only 400 samples, the model cannot diagnose the severity of CKD.

Ping Li et al (09 June 2020) proposed an efficient and accurate system to diagnosis heart disease and the system is based on machine learning techniques. The system is developed based on classification algorithms includes Support vector machine, Logistic regression, Artificial neural network, K-nearest neighbor, Naïve bays, and Decision tree while standard features selection algorithms have been used.

| METHODS | METHODOLGY | CONTRIBUTION | ACCURACY | ISSUES |
|---|---|---|---|---|
| **Predicting Heart Disease at Early Stages using Machine Learning** *Rahul katarya et al.* 04 august 2020. | artificial neural network, decision tree (DT), random forest (RF), support vector machine (SVM), naïve Bayes and k-nearest neighbour algorithm | ML techniques to predict the heart disease at early stage. | 80% of accuracy | Not Effective |
| **Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis** *Noreen fatima, Li Liu et al* 14 August 2020. | 1) **Decision Tree (DT), Naïve bayes algorithm and k-nearest neighbor.** | Article presents the comparative analysis of machine learning, deep learning and data mining techniques being used for the prediction of breast cancer. | Article has 75% accuracy of prediction | They get some knowledge about the major types, symptoms and causes of breast cancer they did not give large issues symptoms of this model. |
| **Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare** *Jian ping Li,Amin Ui haq,salah ud Din et al* 09 June 2020. | Support vector machine, Logistic regression, Artificial neural network, K-nearest neighbor, Naïve bays, and Decision tree | In this article, we proposed an efficient and accurate system to diagnosis heart disease and the system is based on machine learning techniques. | The suggested diagnosis system (FCMIM-SVM) achieved 81% accuracy as compared to previously proposed methods | All FS algorithms results show that the feature Fasting blood sugar (FBS) is not a suitable heart disease diagnosis. |
| **Diabetes prediction using different machine learning approaches** *Priyanka Sonar Prof.K.JayaMalini et al* 29 August 2019. | Decision Tree, ANN, Naive Bayes and SVM algorithms. | The aim of this analysis is to develop a system which might predict the diabetic risk level of a patient with a better accuracy | For Decision Tree, the models give precisions of 85%, for Naive Bayes 77% and 77.3% for Support Vector Machine. | The drawback of the SVM algorithm is that to achieve the best classification results for any given problem and ANN is difficult with dealing with big data with complex model. Require huge processing time. |
| **Deep Learning for Classification and Localization of COVID-19 Markers in Point-of-Care Lung Ultrasound** *Subhankar roy,willi menapace et al* 14 May 2020. | Deep Learning | This paper advances the state of the art in the automatic analysis of LUS images for supporting medical personnel in the diagnosis of COVID 19 related pathologies in many directions. | The acceptable prediction accuracy is 82.7%. | No subsequent testing was done, resulting in the possible inclusion of false positive cases. |

| | | | | |
|---|---|---|---|---|
| **Prediction of Chronic Kidney Disease - A Machine Learning Perspective** *Pankaj chittora,Sandeep chaurasia et al* 22 January 2021 | artificial neural network (ANN), C5.0, logistic regression, linear support vector machine (LSVM), K-nearest neighbors (KNN) and random tree | In this paper using machine learning classifier algorithms, the doctor can detect kidney the disease on time. | It gave the highest accuracy of 98.46%. | The drawback is logistic and KNN algorithm which did not give suitable results. |
| **Fuzzy Rule-Based System for Predicting Daily Case in COVID-19 Outbreak** *Pinar cihan* 17 November 2020 | fuzzy rule basing system (FRGS) | In this study, fuzzy rule basing system (FRGS) used to predict the number of Covid-19 daily cases. | Data with samples of 86 days (around 70% of the dataset) | Due to rapid increase in covid cases sometimes it collapsed to predict the forecasting the patient. |
| **COVID-19 Detection in X-ray Images using CNN Algorithm** *Areej A.wahab Ahmed Musleh et al.* 20 January 2021 | convolutional neural networks (CNN) algorithm | In this paper, to detect covid -19 positive patients using X-ray images. | The acceptable prediction accuracy is 89.7%. | CNN algorithm is used to identify the patient little bit slowly. |
| **Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model** *Vishan kumar gupta,avdhesh gupta et al* 01 January 2021 | Decision tree, Random forest (random Forest),Multinomial logistic regression (multinome),Support vector machine | In this paper they are detecting the COVID-19 cases, that is confirmed, death, and cured cases in India only | If prediction confirmed, death, and cured cases on the testing dataset, random forest has the highest accuracy of 83.54%, 72.79%, and 81.27% on confirmed, death, and cured cases. | It was very difficult to fix an accurate hyperplane between the given values of the dataset. |
| **COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings** *Jordi laguarta et al* 29 September 2020 | Mel Frequency Cepstral Coefficient, Convolutional Neural Network (CNN) | In this paper, they find the covid -19 diagnosis using patient cough audio records. | The model achieves COVID-19 sensitivity of 94.5% with a specificity of 94.2%. | Due to different age groups and regions it is difficult for the metadata to be captured whether cough segmentation can improve the results. |
| **COVID-19 Future Forecasting Using Supervised Machine Learning Models** *Furqan rustam,aijaz ahmad reshi et al* 25 May 2020. | linear regression, least absolute shrinkage and selection operator, support vector machine, and exponential smoothing. | This study aims to provide an early forecast model for the spread of novel coronavirus. | They achieve the 79 percent of accuracy. | It is difficult to store the large number of database. |

| | | | | |
|---|---|---|---|---|
| **Weakly-Supervised Network for Detection of COVID-19 in Chest CT Scans** *Ahmed Mohaammed,congcong wang et al* <u>21 August 2020.</u> | weakly-supervised and Convolution Neural Network | In this paper, we propose an end-to-end weakly-supervised COVID-19 detection. | The accuracy is 76.7% precision and 78.8% F1 score. | The lung part containing severe pathologies, the binarizing method may fail to segment the lung part. |
| **DL-CRC: Deep Learning-Based Chest Radiograph Classification for COVID-19 Detection** *Sadman sakib,tahrat tazrin et al* <u>18 September 2020.</u> | deep learning-based chest radiograph classification (DL-CRC) and convolution neural network (CNN) | In this paper, they propose a viable and efficient deep learning-based chest radiograph classification (DL-CRC) framework to distinguish the COVID-19 cases with high accuracy from other abnormal (e.g., pneumonia) and normal cases | which achieves COVID-19 detection accuracy of 93.94% compared to 54.55% for the scenario without data augmentation. | Due to the lack of labeled data, in this work, our model could not be used to classify the various stages of COVID-19 such as asymptomatic, mild, high and severe. |
| **Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset** *L. J.MuhammadEbrahm* *A. Algehyne et al* <u>5November 2020.</u> | Naïve Bayes Algorithm, Logistic Regression Algorithm, Decision Tree Algorithm, Artifcial Neural Network (ANN), Support Vector Machine. epidemiology labeled dataset. | In this work, supervised ML techniques are used to develop predictive models for the COVID-19 infection, using an epidemiology labeled dataset for positive and negative COVID-19 cases in Mexico. | The result of the performance evaluation of the models showed that decision tree model has the highest accuracy of 94.99% while the Support Vector Machine Model has the highest sensitivity of 93.34% and Naïve Bayes Model has the highest specifcity of 94.30%. | It shows the percentage of COVID-19 positive patients correctly by the models and covid negative patients did not identified accurately. |
| **H3DNN: 3D Deep Learning Based Detection of COVID-19 Virus using Lungs Computed Tomography** *Abdullah aman khan et al* <u>15 January 2021.</u> | 3D Deep learning based methods. | In this paper, They propose an automated 3D deep learning model (H3DNN) to classify 3D chest CT imaging to screen out infected patients. | Hybrid 3D deep learning model achieve 82% of accuracy. | expensive because of 3D deep learning. |

| | | | | |
|---|---|---|---|---|
| **Automatic detection of COVID-19 infection using chest X-ray images through transfer learning** *Elene firmeza ohata et al* 24 September 2020. | k-Nearest Neighbor, Bayes, Random Forest, multilayer perceptron (MLP), and support vector machine (SVM) | In this paper, we propose an automatic system to classify chest X-ray images as from COVID-19 patients or healthy patients using transfer learning with convolution neural networks (CNNs). | Achieving an accuracy and an F1-score of 95.6%. | It does not replace a medical diagnosis since a more through investigation could be done with a larger dataset. |

## 5. DISCUSSIONS:

The existing systems have not delivered their 100 percent accuracy from any outputs taken. They have given only 70 and 80 percent accuracy of output, so with these data the output will not be correct. The existing system is so expensive, if large data is given, the output may not be efficient, different age groups and different groups and different regions metadata are difficult to capture, No subsequent testing was done, resulting in possible inclusion of false positive cases. For detecting the result it takes much amount of time. Using KNN algorithm result is not suitable. But in CNN algorithm it identifies slowly.

So, in proposed system we have better performance of 100 percent accuracy with minimum amount of time.

The proposed system acts as a decision support system and will prove to be an aid for the physicians with the diagnosis. The algorithm, uses various test parameters and makes use of features and data points to predict the relativity of an attribute. The training data is trained by using proposed machine learning algorithm yolo classification clustering.

## 6. CONCLUSION:

Proposed, is a YOLO algorithm for predicting the risk of Covid-19 in a patient using the attributes collected from the Various Tests. Proper adaptation of Data clusters the data into an optimum number of clusters and aids with detecting the normal and abnormal cases efficiently.

## 7. REFERENCES:

[1] C. Huang et al., "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," THE LANCET, vol. 395, no. 10223, pp. 497–506, 2020.

[2] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X. Zhang, and L. Zhang, "Epidemiological and clinicalcal characteristics of 99 cases of 2019 novel coronavirusrus pneumonia in Wuhan, China: A descriptive study," Lancet, vol. 395, no. 10223, pp. 507–513, Feb. 2020.

[3] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, and Y. Zhao, "Clinical characteristics of 138 hospitalized patients 2019 novel coronavirus–infected pneumonia in Wuhan, Jama, vol. 323, no. 11, pp. 1061–1069, 2020.

[4] K. Liu, Y.-Y. Fang, Y. Deng, W. Liu, M.-F. Wang, J.-P. Ma, W. Xiao, Y.-N. Wang, M.-H. Zhong, C.-H. Li, G.-C. Li, and H.-G. Liu, "Clinical characteristics of novel coronavirus cases in tertiary hospitals in Hubei province," Chin. Med. J., vol. 133, no. 9, pp. 1025–1031, May 2020.

[5] T. Guo et al., "Cardiovascular implications of fatal outcomes of patients with coronavirus disease 2019 (COVID-19)," JAMA Cardiol., early access, Mar.27, 2020, doi: 10.1001/jamacardio.2020.1017.

[6] P. Hamet and J. Tremblay, "Artificial intelligence in medicine," Metabolism, vol. 69, pp. S36–S40, Apr. 2017.

[7] M. Jamshidi, A. Lalbakhsh, S. Lotfi, H. Siahkamari, B. Mohamadzade, and J. Jalilian, "A neuro-based approach to designing a Wilkinson power divider," Int. J. RF Microw. Comput.-Aided Eng., vol. 30, no. 3, Mar. 2020, Art. no. e22091.

[8] M. Jamshidi, A. Lalbakhsh, B. Mohamadzade, H. Siahkamari, and S. M. H. Mousavi, "A novel neural-based approach for design of microstrip filters," AEU-Int. J. Electron. Commun., vol. 110, Oct. 2019, Art. no. 152847.

[9] M. B. Jamshidi, N. Alibeigi, A. Lalbakhsh, and S. Roshani, ''An ANFIS approach to modeling a small satellite power source of NASA,'' in Proc. IEEE 16th Int. Conf. Netw., Sens. Control (ICNSC), May 2019, pp. 459–464.

[10] Y. Mintz and R. Brodie, ''Introduction to artificial intelligence in medicine,'' Minimally Invasive Therapy Allied Technol., vol. 28, no., pp. 73–81, 2019.

[11] R. B. Parikh, Z. Obermeyer, and A. S. Navathe, ''Regulation of predictive analytics in medicine,'' Science, vol. 363, no. 6429, pp. 810–812, Feb. 2019. [12] A. Becker, ''Artificial intelligence in medicine: What is it doing for us today?'' Health Policy Technol., vol. 8, no. 2, pp. 198–205, Jun. 2019.

[13] N. J. Schork, ''Artificial intelligence and personalized medicine,'' in Precision Medicine in Cancer Therapy. Cham, Switzerland: Springer, 2019, pp. 265–283.

[14] M. B. Jamshidi, M. Gorjiankhanzad, A. Lalbakhsh, and S. Roshani, ''A novel multiobjective approach for detecting money laundering with a neuro-futurezy technique,'' in Proc. IEEE 16th Int. Conf. Netw., Sens. Control (ICNSC), May 2019, pp. 454–458.

[15] M. B. Jamshidi, N. Alibeigi, N. Rabbani, B. Oryani, and A. Lalbakhsh, ''Artificial neural networks: A powerful tool for cognitive science,'' in Proc. IEEE 9th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON), Nov. 2018, pp. 674–679.

[16] S. Huang, J. Yang, S. Fong, and Q. Zhao, ''Mining prognosis index of brain metastases using artificial intelligence,'' Cancers, vol. 11, no. 8, p. 1140, Aug. 2019.

[17] P. Rashidi and A. Bihorac, ''Artificial intelligence approaches to improve kidney care,'' Nature Rev. Nephrol., vol. 16, no. 2, pp. 71–72, Feb. 2020.

[18] D. S. W. Ting, L. Carin, V. Dzau, and T. Y. Wong, ''Digital technology and COVID-19,''Nature Med.vol. 26, no. 4, pp. 459–461, 2020.