A Novel Ensemble-based Classifier for Detecting the COVID-19 Disease for Infected Patients

Prabh Deep Singh[1] · Rajbir Kaur[2] · Kiran Deep Singh[3] · Gaurav Dhiman[4]

Abstract

The recently discovered coronavirus, SARS-CoV-2, which was detected in Wuhan, China, has spread worldwide and is still being studied at the end of 2019. Detection of COVID-19 at an early stage is essential to provide adequate healthcare to affected patients and protect the uninfected community. This paper aims to design and develop a novel ensemble-based classifier to predict COVID-19 cases at a very early stage so that appropriate action can be taken by patients, doctors, health organizations, and the government. In this paper, a synthetic dataset of COVID-19 is generated by a dataset generation algorithm. A novel ensemble-based classifier of machine learning is employed on the COVID-19 dataset to predict the disease. A convex hull-based approach is also applied to the data to improve the proposed novel, ensemble-based classifier's accuracy and speed. The model is designed and developed through the python programming language and compares with the most popular classifier, i.e., Decision Tree, ID3, and support vector machine. The results indicate that the proposed novel classifier provides a more significant precision, kappa static, root means a square error, recall, F-measure, and accuracy.

Keywords Machine learning · COVID-19 · Artificial intelligence · Quality of service · Corona virus · Ensemble classifier

1 Introduction

The SARS-CoV-2 that causes COVID-19, colloquially known as coronavirus, is a contagious viral in the coronavirus family. However, the virus genome sequence analysis indicates

✉ Gaurav Dhiman
gdhiman0001@gmail.com

Prabh Deep Singh
ssingh.prabhdeep@gmail.com

Rajbir Kaur
rajbir277@yahoo.in

Kiran Deep Singh
kdkirandeep@gmail.com

1 Department of Computer Science & Engineering, Punjabi University, Patiala, Punjab, India

2 Department of Electronics & Communication Engineering, Punjabi University, Patiala, Punjab, India

3 Department of Computer Science & Engineering, IKG Punjab Technical University, Punjab, India

4 Department of Computer Science, Government Bikram College of Commerce, Punjabi University, Patiala, Punjab, India

that this belongs to the community of coronavirus β-CoV generations. The novel coronavirus was confirmed to the World Health Organization on 31 December 2019 and occurred in Wuhan, and has travelled worldwide. The virus developed a global threat, and on 11 February 2020 was called COVID-19 by W.H.O. It is the family of SARS, ARDS, which has been listed as a public health emergency (Baudier et al. 2021). While a healthy person is in contact with an infectious person, the virus is spread through the respiratory tract (Benreguia et al. 2020). The virus transmits by air and physical contact and penetrates respiratory cells by binding to the enzyme that transforms angiotensin 2. The most frequent infection symptoms are breathlessness, fever, cough, smelling and taste loss, headache, and muscle disease. As the virus evolves, it causes several issues in all facets of human life, and new problems occur with time. New strategies are being developed every day to address these rapidly evolving problems.

The health sector is now at the frontline of the epidemic. Healthcare is increasingly challenging to manage because of insufficient and less effective care that does not meet the increased demands of a growing ageing population with chronic diseases (Farahani et al. 2018). It is an extensive industry that needs to collect and process medical data in real-time. This sector's critical issue is handling evidence that

needs in-patient prevention and distribution of knowledge to clinicians for fast-track medical treatment. The sector has sought together, handle, and restore knowledge to strengthen patient procedures and technology advancement by leading actions such as surgeons, sales associates, hospitals, and healthcare firms. The protection of patients' health records, access control for regular and emergency cases, and smart data deduplication to conserve storage space in large data storage systems (Yang et al. 2019). However, the management of health data has lately become a complex challenge due to the immense amount of data, security vulnerabilities, the incompetence of cellular network applications, and the speed it increases.

Diseases such as COVID-19 are the greatest concern of all nations, particularly the healthcare team and their families (Abdel-Basset et al. 2021; Dhiman et al. 2021). There is no evidence of any subsequent change the next day. Due to the immediate need for research with COVID-19, new technology is expected to be built to provide quick

and efficient analysis. These primary innovations include artificial intelligence (AI) (Kaur et al. 2021; Dhiman and Kaur 2019c, Dhiman and Kaur 2020a, c, 2020), big data, the Internet of Things (IoT), the Internet of Medical Things (IoMT), 5G, and Blockchain to accomplish clear objectives (Kaur et al. 2020; Dhiman 2019a, b, d, 2020b). In order to address the questions about what are disruptive technologies and using them to reduce the COVID-19 outbreaks, Table 1 provides examples of useful disruptive technologies which can better be used to limit the COVID-19 outbreaks (Lasi et al. 2014; Chen et al. 2020; Chamola et al. 2020; Wang et al. 2018; Singh et al. 2020c).

AI and machine learning are useful for virtually all decision-making processes. Both screening methods boost the precision of diagnosis for infections, both contagious and non-infectious. Artificial intelligence and machine learning researchers have been eagerly looking and waiting for real-time data produced by this pandemic worldwide. Consequently, timely distribution of COVID-19 patient

Table 1 The disruptive technologies and their advantages to limit the COVID-19 outbreaks

Disruptive Technology	Description	Advantages
Internet of Things (Gupta et al. 2018)	A key part of the Fourth Industrial Revolution, connects services and semantics via wireless protocols	IoT is useful to avoid COVID-19 outbreaks. Sensors will regularly follow up hospitalised patients or home quarantine patients. IoT will ensure the patient is being taken care of appropriately. The IoT can be used to screen individuals with COVID-19.
Industry 4.0 (You and Feng 2020)	Industrial 4.0 used the IoT to use resilient services.	Industry 4.0 will be used to check up on the first vital signs of hospitalized patients.
The Internet of Medical Things (Cao et al. 2019)	IoMT is a new healthcare technology approach that incorporates data-driven information technology and contemporary medicine techniques.	IoMT usage can help patients' physical and psychological healthcare in the hospital or those under home quarantine due to trust and protection.
Virtual Reality (Sood and Singh 2018)	Virtual Reality technology is used in a computer programme to simulate an environment.	The video call is appropriate for patients who are in hospital or quarantine.
Artificial Intelligence (Singh and Kaur 2020b)	Artificial Intelligence is a powerful tool that lets computers learn and think.	It has developed robots to assist in the medical examination of patients.
Big Data (He et al. 2018)	Big data is a mechanism that gathers and processes vast quantities of information that programmes cannot typically process.	Big data analysis can generate several statistics about COVID-19 outbreaks and other reports to minimize the pandemic's impact.
5G (Singh and Sood 2020a)	The fifth-generation technology can support global mobile networks	The 5 G would provide the healthcare team with high-paced communication to enhance the track of COVID-19 control and review patient cases.
Blockchain (Nair et al. 2020)	Blockchain is a transaction record between two parties	The Blockchain can be used to restrict the influenza pandemic by combining various data sources.
Drone technology and Autonomous Robots (Angurala et al. 2020)	autonomous robot/Drones can carry out specific responsibilities without the intervention of external agencies	The robot takes over many tasks, such as medical care monitoring and supervision, from the team of health professionals.

data, such as physiological characteristics and clinical outcome of COVID-19 patients, accompanied by corresponding data translation for easy access, is highly significant yet challenging (Abdel-Basset et al. 2021). The link between health care and physicians derives from the first AI expert framework that serves as clinical decision support for physicians and medical experts. Researchers also found proof that AI and ML may diagnose and monitor different diseases, from pandemic to non-communicable diseases. The continuous developments in AI and ML have dramatically enhanced therapy, medication, screening, estimation, forecasting, touch tracing, and drug/vaccine production method for the Covid-19 pandemic, thus reducing human involvement in medical practice.

AI technologies improve the performance and precision of their activities. The healthcare sector desperately needs decisions to contain the virus to help them get good advice in real-time to prevent the spread of the virus (Small and Cavanagh 2020). AI can professionally emulate human intellect. It is used to scan, evaluate, forecast, and monitor existing patients and potentially potential patients accurately. Significant applications for monitoring confirmed, rescued, and deadly data are applicable. Hospitals, public health departments, and private health businesses search for open means of screening patients for COVID-19 symptoms, such as online symptom checkers.

COVID-19 is spread around the world, each bit of creativity in science and imagination is taking us closer to solving this pandemic (Abdel-Basset et al. 2020c). Understanding and solving the COVID-19 problem, artificial intelligence (AI), and machine learning play a crucial role (Xu et al. 2020). The research and the development of methods imitating human intelligence is artificial intelligence (Latif et al. 2020). It is very optimistic that Artificial Intelligence will become an exciting field of study in the face of the present obstacles as a result of its achievements in the fields of diagnosis, care, patient management, medicine discovery, Epidemiology,

etc. Machine learning technology helps machines imitate human intelligence and absorb vast data quantities to find trends and observations quickly (Qureshi et al. 2017). Further ML and AI algorithms enable medical treatment, and follow-up plans to be diagnosed and optimized to achieve improved outcomes (Ren et al. 2020). A powerful platform for gathering data and real-time contact for all health professionals who battle COVID-19 on the frontlines allows them to track diagnosis and keep patient knowledge updated in real-time. Progress in technology has fast implications in any area of life, whether in medicine or elsewhere. Through taking decisions by evaluating the evidence, artificial intelligence has shown positive results for health care. COVID-19 hit about 100 countries in no time. Citizens around the world will, in the future, be vulnerable to its impact. A control scheme for the detection of the coronavirus should be established. The identification of disease using different AI methods may be one of the alternatives to the existing havoc regulation.

A new and valuable method to diagnose early coronavirus infections and track infected patients' status is artificial intelligence. Artificial intelligence By designing practical algorithms can significantly improve medication quality and decision making. The necessary process of diagnosing the disease at an early stage is shown in Fig. 1. AI is useful not only in the care of compromised patients Covid-19 but also in the proper control of their health. Since the beginning of the epidemic, the use of AI approached drug production. AI has been widely used in studies in the development of new molecules. AI will create an intelligent network to track and forecast the propagation of this virus automatically. To obtain the visual features of this condition, a neural network may also be created, which will help to track and manage the individuals affected. It can provide regular patient feedback and also to provide remedies for the COVID-19 pandemic to be followed. Nurses can view patient records from anywhere safely through flexible and compact computers. The wireless feature helps telehealth

Fig. 1 Detection of disease at early stage
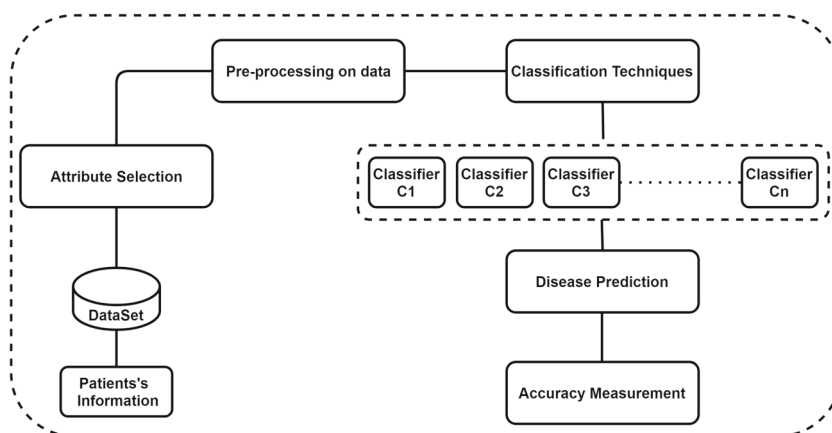
or telemedicine implementations as nurses can consult, log, and exchange information with the physician. This would increase productivity and save time for routine inspections.

The use of machine-learning chatbots for contactless COVID-19 screening and answering questions from the public are now part of healthcare and government agencies. Machine learning also helps researchers and experts evaluate vast quantities of data for forecasting COVID-19 distribution so that the potential pandemics can be detected as an early warning mechanism and endangered populations can be identified (Sharma et al. 2020). The amount of knowledge on COVID-19 is exponentially rising in healthcare practitioners and researchers and making it impossible to obtain perspectives that inform treatment.

Ensemble-based classifiers motivated us to propose a novel classifier for detecting the COVID-19 disease for infected patients, which forecasts the spread of epidemic transmission. It is suitable for applications that need precise classifications, real-time responses, and short reaction times. Medical machine learning classification predictions were made in the past. Furthermore, various algorithms for classification are required for the same type of domain. These algorithms' output is based on the non-serious, monotonous, and redundant properties of the data. Two methods are also suggested: one to increase the precision of the proposed novel ensemble-based classifier and another to minimize the proposed novel classifier's speed by reducing the number of data samples.

This paper's main contribution is to encode, outside of the conventional healthcare system, smart health care, and travel information using machine learning algorithms to detect an infected COVID. This work will encourage researchers to work further to build a solution to help predict COVID-19 outcomes for patient health, which incorporates patient population, moving, and subjective health data.

The rest of the paper is organized as Section 2 provide the literature review; the methodology is explained in Section 3. Section 4, experiment setup and results are shown, Section 5 comes with discussion, and finally, the concluding remarks given by Section 6.

2 Literature Review

The transformative technology of 21st century artificial intelligence (AI) has seen numerous uses in weather forecasts and astronomical analysis to autonomous systems. We observe some similar work in which AI was tried to track, deter, and forecast the pandemic COVID-19. Several scientific study papers were examined regarding Artificial Intelligence (AI) application to the coronavirus response. The majority of writers concentrated on a critical evaluation

of models to forecast the probability of disease growth, hospital acceptance, and progression. It is hard to keep up with what is going on in health care due to the constant appearance of new diseases, and viruses (Chang 2018c). However, most epidemiological studies aimed to predict and prevent the COVID-19 with better accuracy were excluded.

Gomathi et al. (2020) provide an alternative for the creation of a mechanical prognostic model that precisely anticipates the survival of individual severe patients with more than 95% of clinical evidence from various sources. This paper aims to include a detailed analysis of the diagnostic aspect of COVID-19, update the cost-effectiveness and timeliness of criticism, and predetermined. Marmarelis (2020) suggest a new model-based for predicting the total cases of COVID-19 using US data. Machine learning is used to design and develop the method. Ouyang et al. (2020) proposed a new CNN web treatment module to focus on lung infection areas in diagnostic decisions, which is directly related to COVID-19. The sizes of the infection areas between COVID-19 and CAP are unbalanced in dissemination, partly due to the rapid development of COVID-19 after symptoms began. Therefore, a double sampling approach is being developed to minimize imbalances. Ahn et al. (2020) concentrate on South Korea's epidemiological study methods after a brief comparison of contextual discrepancies with French countries. Authors analyze the use habits of original data, de-identification data, and encrypted data to assess personal privacy and health concerns. Also, They address the COVID-19 index, which includes collective illness, the severity of outbreaks, the availability of health care services, and the death rate. Li et al. (2020) propose to use a limited number of COVID-19 CT exams and an archive of negative samples for successful and productive COVID-19 grade training networks. The idea is to remove COVID-19 features and negative samples using a new self-monitored learning approach in concrete terms. A pre-set number of negative samples is chosen and supplied to the neural network. Chamola et al. (2020) analyzed the issues related to the COVID-19 pandemic by many credible outlets. Their research highlighted its effect on the international economy and the direct health effects of the outbreak of COVID-19. The authors also investigate the use of technology such as the IoT, UAVs, AI, and Blockchain (AI) to mitigate the consequences of the COVID-19 epidemic. Abdel-Basset et al. (2020a) presented a groundbreaking semi-supervised segmentation method that worked well with only a few CT scans. They developed a novel dual-path deep-learning architecture to avoid channel information loss without losing any high-level knowledge. This contribution overcomes the lack of availability of CT scans by obtaining a large number of scans. The study provided insight into how objective lung disease diagnosis can be made in

small data environments. Ulhaq et al. (2020) analyzed the impact of COVID-19 in the market and propose a solution by computer vision. The author collects the data from various sources like the WHO's official website and used various research tools to analyze. Abdel-Basset et al. (2020b) presented a hybrid image segmentation method to overcome the image segmentation problem using the thresholding technique with Kapur's entropy maximization algorithm. The results show that the proposed algorithm is better than the existing algorithms. Li (2020) designed a hierarchical two-tier combined approach to speed up the asymptomatic screening of COVID-19. The first stage divides a population into groups, resulting in the acceleration of groups. At the second stage, a group is sub-grouped, leading to an intra and inter-subgroup acceleration. Using this computational method and numeric algorithms, optimum Group size is calculated. Niu et al. (2020) explain and reflect the dissemination of COVID-19 and propose a paradigm for human migration, where "exposed" individuals may be infectious, for sensitivities-exposed-infested-restricted. The disease's primary reproductive number and its connection with the model parameters are derived from this model. Jamshidi et al. (2020) provided an answer to artificial intelligence battling the virus. A variety of deep learning approaches, including GNE, Intense Learning Machines, and Long-Term Short Memory, have been demonstrated (LSTM) to accomplish this goal. It describes an applied bioinformatics methodology in which numerous knowledge aspects shape the user-friendly platforms for physicians and researchers from a continuum of organized or unstructured data sources. Quatieri et al. (2020) diagnosed and tracked COVID-19 through asymptomatic or symptomatic stages, and the Authors suggest a voice making and signal processing system. Methods: The approach is based on the dynamics of neuromotor synchronization through respiratory expression, phonation, and articulation subsystems based on COVID-19's features of lower and upper inflammation of the respiratory system and growing evidence of neurothral viruses. Wang et al. (2020b) proposed a new method for joint learning to reliably classify the COVID-19 through efficient learning of heterogeneous, dispersed data sets. They create a clear context through the recent redesign of the COVID-Net, which will boost predictive precision and performance in network architecture and learning strategy. Wang et al. (2020a) provide COVID-19 screening in 3D chest images. The system will predict accurately whether or not a CT scan involves pneumonia while distinguishing pneumonia forms caused by other viruses between COVID-19 and Interstitial Lung Disease. Two 3D ResNets in the proposed approach are connected into one model for the above two tasks utilizing a new technique for priority

attention. Chang (2018b) presents the new research in data processing and simulation for bioinformatics and healthcare. The training data with a low completion time to achieve malignant tumour and genes simulations to inspect their status and querying the output data within seconds. The malignant tumour and gene simulation can achieve 360 degrees for an inspection of cancerous presence.

Chang (2018a) Showed how computer science could model medical imagery to investigate regions, including genes and protein-based simulations of cancer growth and immunity, that are not readily attainable. The concept is similar to the digital surface theories to simulate how biological units can form larger units until the whole biological subject unit is created.

There are significant pancreatic cancer clinical signs. Because it is challenging to diagnose pancreatic cancer, every one symptom is not a good cancer predictor. Various combinations and individual contributions to these functions are also considered. Class 0 is 90 per cent (Non-pancreatic cancer), and class 1 is 10 percent.

3 Methodology

The data collection of COVID-19 may be obtained from different sources such as a log, database, sensor, and several other sources, but the data must be clean and well ordered to conduct the analysis method. Therefore, extracting the data would have the job done. An artificial synthetic data for COVID-19 is synthesized. Class Restoration seeks to restore COVID-19 into a modern, better-performing version. To more efficiently process the COVID-19 results, the data is decomposed such that the machine learning algorithm can provide more manageable processing. Standardization means the functions have the characteristics of normal zeros and the usual standard deviations of one. Each experiment sample is separated into a training sample of 67% and a research sample of 33%. Randomly sample is taken, set hyperparameter, trained the model on it, and then evaluated how well the model worked. The data collection of COVID-19 is incredibly unbalanced. The algorithm defaults towards the COVID-19 positive cases if we use this dataset to train our model, which eventually results in the model's lousy output being tested on invisible data. The theory of ensemble modelling is to integrate multiple models to predict the result of a variable. A consensus that ensemble models will improve predictive precision is commonly held. We have the hyperparameters set in each device and have passed the reequipped training set as training data for each model. These models have been tested with different parameters, including Precision, Kappa Static, Root Mean Square Defect, Recall, F-measure, and accuracy.

3.1 Data description

In machine learning, data sets are an important component. Data collection is one of the most difficult activities, especially in connection with the clinical domain as early detection. Table 2 shows the data type used for different symptoms of COVID-19 patients.

Our dataset includes the symptom base COVID-19 positive and COVID-19 negative cases. The dataset contains 824,548 cases, out of which only 1376 are COVID-19 positive. As the optimistic class accounts for just 0.167% of the overall COVID-19 data, the dataset is rather unequalled. The data collection comprises numerical values resulting from the transformation of the principal component analysis (PCA). However, the initial features were not released because of the secrecy issue. There are 18 characteristics in all, of which the principal component analysis produced 16 features. PCA is a method for minimizing dimensionality, in which several initial variables condense into a smaller subset of function variables. The principal component analysis (PCA) is to minimize the dimensionality of a data package composed of multiple associated variables to the greatest degree while preserving the variance present in the dataset. The same may be achieved by converting the variables into a new collection of variables known as the main components and orthogonal to reduce the uncertainty preserved in the initial variables as we step down. Thus, the maximum variance of the 1st major factor of the initial components persists. The core elements are the prophetic vectors of the orthogonal matrix. Essentially, it is important to scale the dataset to be used in the PCA. The findings are sensitive to relative scaling as well (Fig. 2).

Algorithm 1 Dataset generation algorithm.

Result: COVID-19 Data Creation
Input   : Patient Symptom
Output: Generate COVID-19 datasets

1 Step I: N initialize to 1
2 Step II: for N ¡= number of entries required do
3 Step III: allocate values to significant symptoms of COVID-19 based on their probabilities.
4 Step IV: allocate values to the minor symptom of COVID-19 based on their probabilities.
5 Step V: create new user data by combining Step III and step IV values generated.
6 **if** $PatientID_{Result} == COVID\text{-}19\ Positive$ **then**
7   The same user data is already present then
8   Discard this data
9 **else**
10   Add new user data
11 **end**
12 increment N

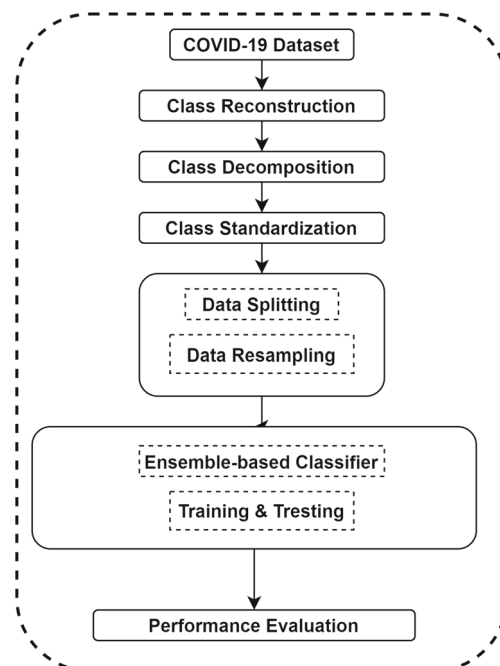**Table 2** Data type for different symptoms of COVID-19 patients

S. No.	Parameter	Data Type
1	Body Temperature	Float
2	Difficulty in breathing	Boolean
8s3	Loss of speech or movement	Boolean
4	Room Temperature	Float
5	Eye colour pink or not	Boolean
6	Age	Int
7	Blue lips/face	Boolean
8	Body ache	Boolean
9	Loss of Taste/smell	Boolean
10	Travel History of last 14 Days	String
11	Diarrhea	Boolean
12	Running nose	Boolean
13	Already suffering from any other disease	Boolean
14	Sore throat	Boolean
15	Tightness of chess	Boolean

Fig. 2 Proposed Methodology for novel classifier

Algorithm 2 Principal component analysis

    Input : COVID-19 data set
    Output: An updated COVID-19 data set

1 Calculate the association matrix of the collection of data points.
2 Calculate eigenvectors and their eigenvalues.
3 Sum the eigenvalues of the original eigenvectors.
4 Choose first, and that will determine the new dimension.
5 Reduce the n individual data points to a k dimensional space.

3.2 Class Reconstruction

Class Reconstruction strives to rebuild the COVID-19 dataset into a form that provides better performance. Specifically, in this paper's case, classes are decomposed using a clustering algorithm; each cluster can then be used to identify a new class label set. The method uses Convex Hulls to attempt to identify clusters that can be assigned the same class label.

A geometrical approach provides a variety of solutions to the problems faced frequently in machine learning. Convex hulls are chosen as the modus operand for finding this pattern. An additional step is added to the user for two reasons: they reduce the size of the data being handled, making the process much faster for larger datasets, and they make it easier to identify interference between clusters.

Figure 3 shows the clusters of the two classes of a sample dataset (green and red) and their respective convex hulls. All further steps to identify superclusters will be performed using these convex hulls. As shown clearly in Fig. 3, the red class clusters can quickly be joined to form

a supercluster. This optimization is identified by finding the red supercluster (as shown in Fig. 4.

The convex hull for this supercluster, "red super hull", is then added to a new dataset along with one point from the convex hull for one of the green cluster convex hulls. A new convex hull is calculated for this dataset to identify if Hull Point is part of the convex hull. If it is, then it must be placed geometrically outside the red super hull, causing no interference. However, if the Hull Point is not part of the newly calculated convex hull, it is sitting in between the two red clusters, preventing a supercluster formation. This step is repeated across all points in the convex hull. This process is then performed on all remaining convex hulls of all non-red class clusters (in this case, all three green clusters must go through the process). If no interference is found, then the red clusters can be joined together to form a supercluster. The exact process is repeated for the green class clusters, and it is confirmed that some interference is found between all pairs of clusters, preventing them from being merged. Early on in experiments, it was identified that there was often very little interference between two clusters preventing a supercluster's formation. So an additional parameter, called "gamma" γ, was added to the process that could help control what level of interference was considered significant. With low gamma values, even a small level of interference will prevent the formation of superclusters. With a higher value of gamma, however, small interferences can be ignored, and a higher number of superclusters can be formed, as shown in Fig. 5

3.3 Data decomposition

Data decomposition is used to reduce the size of a COVID-19 dataset to make the machine learning algorithm's

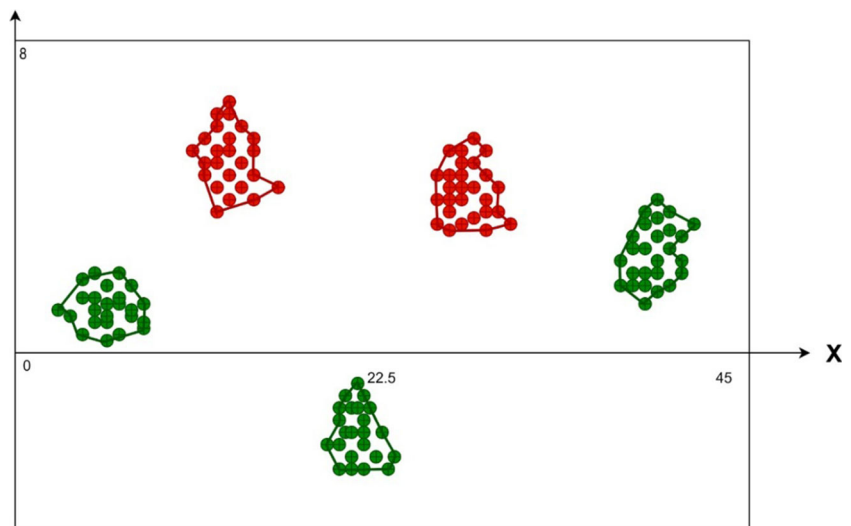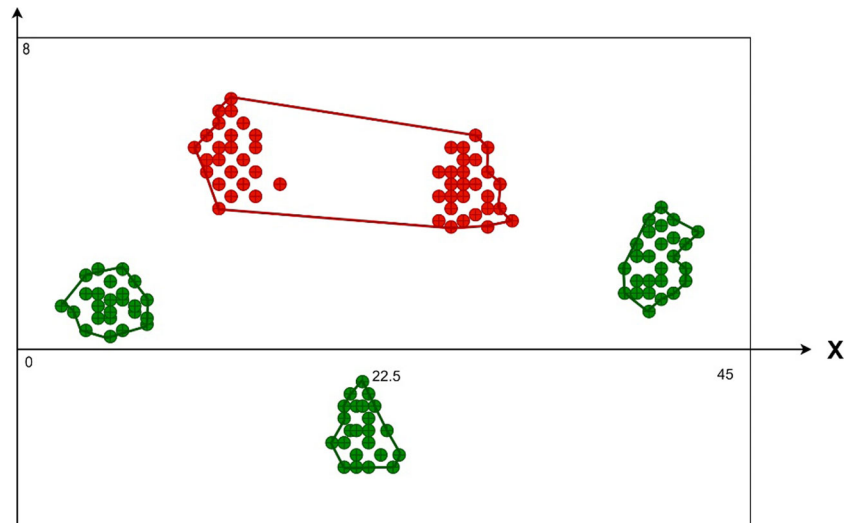**Fig. 3** Convex Hulls of identified clusters by class

Fig. 4 super cluster

processing more manageable. When it comes to the proposed novel ensemble-based classifiers, large datasets need very high-performance computing to manage. The goal is to identify a method to reduce the dataset size without bearing a significant reduction in accuracy (Imran et al. 2020). The process is used clustering algorithms followed by a convex hull-based algorithm to reduce the dataset size by a large percentage; in most cases, the modified dataset size will be less than 15% of the original size. This will help make it easier and faster to analyze the proposed novel ensemble classifier's performance on the modified dataset.

The approach is taken to reduce dataset size attempts to consider each class's geometric nature in the dataset. Datasets that contain classes with significantly dispersed data will especially benefit from a clustering-based algorithm. Since this algorithm is attempting to improve speed, naturally, the type of datasets that will be most applicable for this approach are large ones (Figs. 6, 7, 8 and 9).

Each class in the given dataset is run through the clustering algorithm. For each cluster, the convex hull is identified and added to the modified dataset. The size of this dataset is typically much smaller than the original dataset. Often, for datasets with higher class dispersion, the dataset tends to get bigger. This helps maintain the number of points required in the modified dataset to maintain accuracy from a trained machine learning algorithm. Using the clusters' convex hulls instead of random points is to ensure that the geometric nature of each cluster is captured and to ensure that a good class boundary can be found. Creating a modified dataset with the above steps alone may not be sufficient. An optional step has to be added that will have to be used for certain datasets. If a dataset has vastly different classes in size, the user must ensure that their relative strength in the modified dataset does not change significantly. Essentially, it asks the user to take a cluster and find its convex hull. These points are then added to the
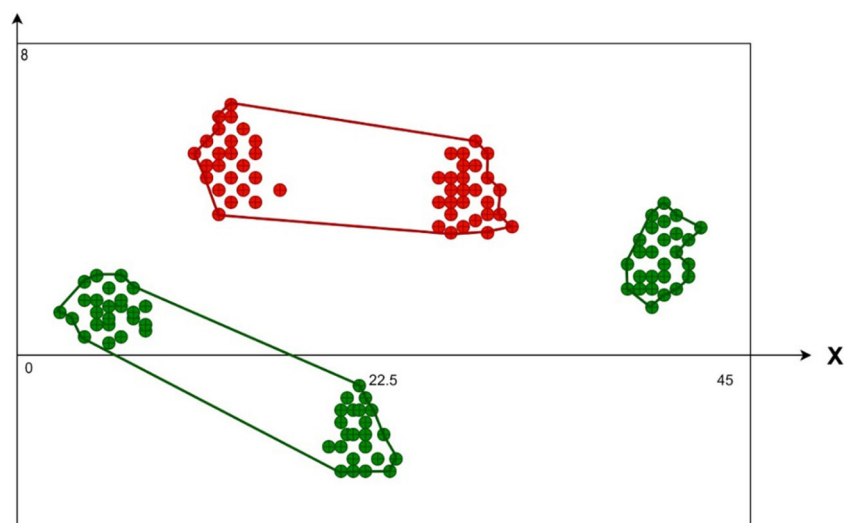
Fig. 5 Two green class clusters
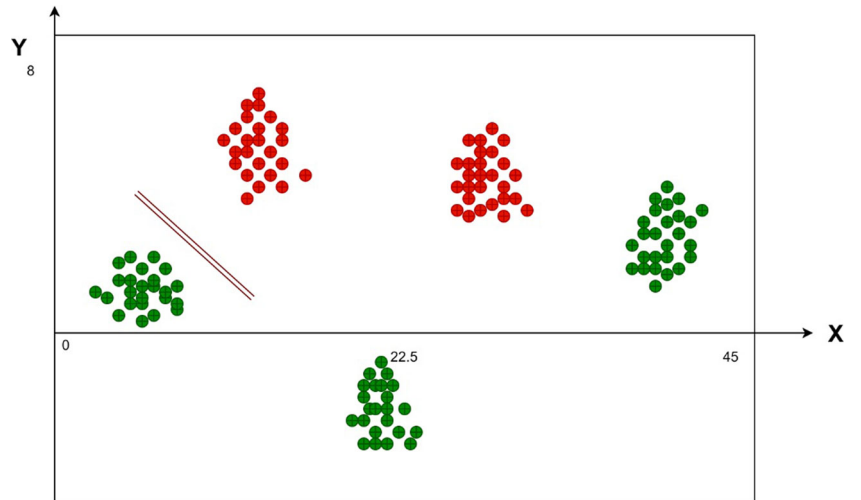
Fig. 6 Dataset with one class
boundary
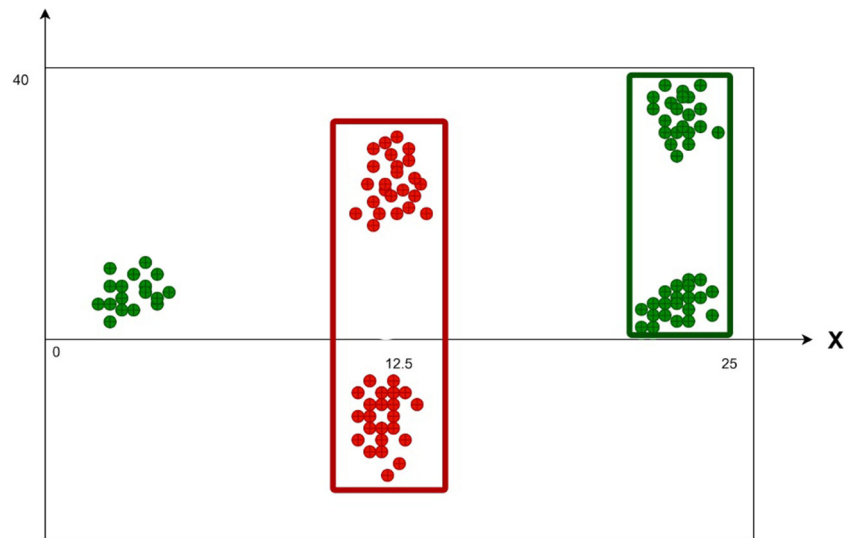

Fig. 7 Super Cluster
combination

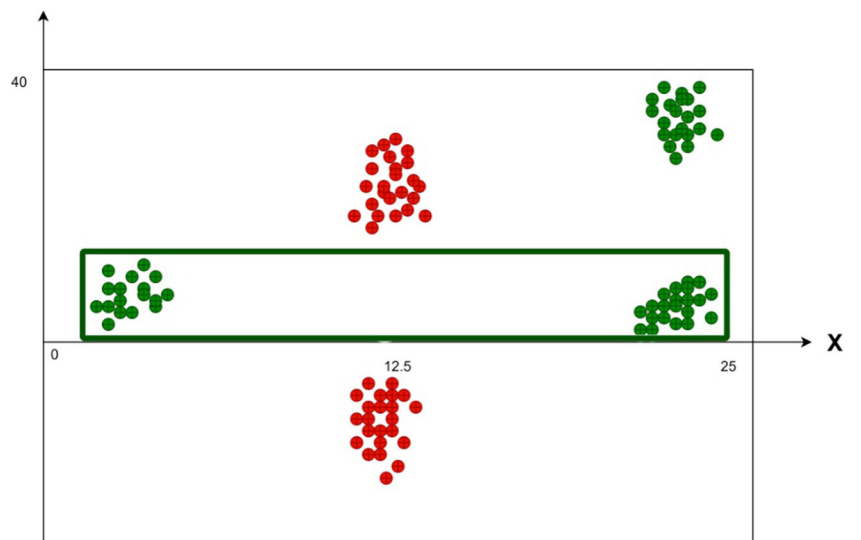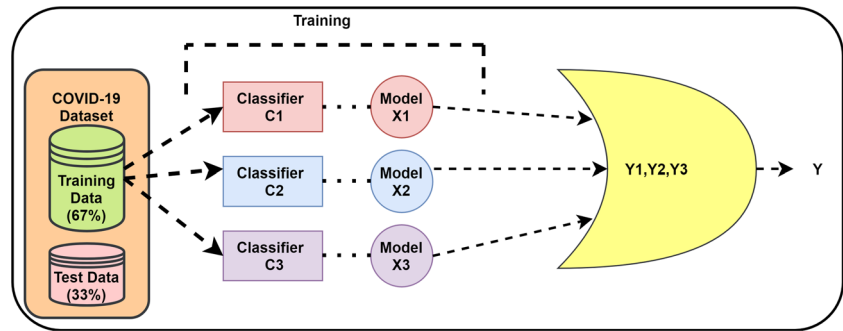**Fig. 8** A better Super Cluster
combination

Fig. 9 Proposed Ensemble
Based Classifier

modified dataset and removed from the original cluster. The
new cluster is rerun through the convex hull calculation.
This second layer convex hull is then also added to the
dataset. This process is continued in parallel for all Class
B clusters until the desired size has been reached for Class
B in the modified dataset. This method's benefit is that
each point selected will strengthen each cluster's geometric
nature in the modified dataset and hence contribute to
a better-learned algorithm. The modified dataset obtained
using the above method is used to train the required machine
learning algorithm. The original dataset is then evaluated
using this trained algorithm, and the accuracy and time
taken recorded. The original dataset is then itself used to
train the required machine learning algorithm. Its accuracy
and time taken to train are also recorded.

3.4 Data standardization

Standardization ensures the functions are re-scaled such
that they have the properties of a standard typical zero
and standard deviations of one. Many machine learning
algorithms used for classification need features to be
standardised (Elhadad et al. 2020). If the standardization is
not carried out, the predicted value is changed.

3.5 Data splitting

We divided the whole dataset into a training set of 67% and
a test set of 33% for each experiment. We used the training
set to resample, set hyperparameter, and train the model,
and we tested how well the learned model is performed. We
specified a random seed when dividing the data, ensuring
that the same data was divided every time the program was
running.

3.6 Data resampling

The data set is too unbalanced. In this case, the algorithm
defaults towards the COVID-19 positive cases if we use this
dataset to train our model, which ultimately results in the
bad performance of the model being evaluated on invisible
data (Pham et al. 2020).

3.7 Hyperparameter tuning using 10-fold cross validation

Hyperparameter is the external model configuration whose
data is not used to determine the model's importance.
The hyperparameter can not be confused with the model
parameter (internal configuration of the model), which is
part of the model and whose value is defined during the
training cycle and can be measured. Since hyperparameter
is outside the formula, it is manually set by the practitioner.
However, the value must be calibrated accurately in advance
to give the most efficient outcomes (Hussain et al. 2020).
Cross-validation is the tool for performing hyper-parameter
tuning. We used K-fold cross-validation where we used K
to be 10.

We use a ten times cross-validation training package,
where each fold consists of 10% of the data. Then, we
analyze the model to decide whether it works well. It is
often used to operate hyperparameters which can be used
to prevent overfitting (De Santis et al. 2020). By running
cross-validation iterations, we found the most substantial
hyperparameter values. As a consequence, when working
with each product, the hunting for each model was quite
comprehensive.

3.8 Ensemble method

Classifying train data by a particular algorithm contributes
to the formulation of a theory used for predictions.
Regarding train data, residual is the difference between
the approximation hypothesis and the actual value, while
the error concerning test data and anonymous data is the
gap between the predictions and the actual values, which
consists of bias, variance and noise. Inference bias occurs
where the model does not correctly match the training
results, showing the researcher that the data was not
adequately trained. A lower bias generates less residual,
which thus less prone to lead to absolute error. Variance
arose from the algorithm overfitting the training sample,
reflecting the volume of noise unintentionally used in the
model. The mistake is distinct from the residual and not
the same. If a model is too responsive to noise, it will

not be well generalized between known and unknown results. The model's bias and variance are created not to be measured; the data's noise is embedded in the data, so it cannot be extracted. It is impossible to eliminate bias and variation concurrently because there is a trade-off in accurately accounting for each of these outlets. If the amount of features in the model rises, the bias reduces while the variance continues growing. In the long run, the difference will decline as the number of samples rises. As the noise gets more complicated, the error reduces first and then rises as the uncertainty develops. To enhance a classifier's efficiency, ensemble methods are used to minimize one or both of the two sources of error by integrating several essential or even "weak" classifiers. The principle of ensemble modelling is to construct a predictive model by incorporating several models. An agreement that ensemble methods can boost overall predictive efficiency is well-known.

A novel Ensemble method is one of the most effective multi-classifier approaches, consisting of combining a set of classifiers of the same type to get a single more efficient model. Nowadays, many methods are automatically capable of generating sets of classifiers. The novel ensemble-based classifier uses the concept of random forests. The procedure is applied to the required degree of precision. The novel approach is based on the active ingredient of decision trees. These classifiers are not incredibly useful separately, but together, they create a strong classifier that is especially stable (Liu and Wang 2010). The random function is that there are only a small amount of trees in the ensemble, and the aim is to build a diverse ensemble. The formalism of random forests arose centred on the usage of decision trees as basic classifiers and the incorporation of randomness in their inference.

The existing methods in state of the art try to deal with randomization. The first contribution is the introduction of a new method to generate a set of classifiers. This method combines Bootstrap Sampling, Random Subspaces, and random forests to generate a more efficient set of trees than each method individually. A novel ensemble-based classifier is proposed. In this paper, we also deal with tree aggregation and ensemble selection in the novel ensemble methods. Classical random forests use a majority voting to aggregate the decision of each classifier. This technique is not optimal since it gives the same weight to each tree's decision even they have not the same performances. We propose a weighted voting mechanism to random forests in our second contribution, which give better results than classical majority voting. The third contribution is a tree selection method in a forest to keep only the best trees. This technique belongs to the family of ensemble selection or pruning methods.

Algorithm 3 Ensemble based classifier.

> **Input** : A new incoming data instance d, a Decision tree
>
> **Output**: An updated Decision tree including d_i

1   Navigate instance d to leaf node starting from the root of the tree;

2   Update statistics on attribute values at current node including instance d ;

3   Evaluate gain function $G(A_i, S)$ on current node for all examples seen so far S.;

4   **if** *There exists i such that for all j, $G(Ai, S) > G(Aj, S)$* **then**

5   **if** *(Training size is less than its threshold and tree depth is less than its threshold* **then**

6   split the node with attribute A_i ;

7   initial statistics for each child node ;

8   **end**

9   **else**

10   Select feasible candidate;

11 **end**

3.9 Training and testing phase

The Test dataset is the data used for assessment. It is only seen once in a single training run. The comparison collection is used for the assessment of rival versions. On many occasions, the validation set is used as a test set, but it is not a safe idea since it may cause messy outcomes. The test range was deliberately chosen for accuracy. The model includes publicly accessible data obtained from several groups that the model would be evaluated. We have set hyperparameters in each model and transferred the reequipped training set as training data on each model after tuning the hyperparameters on each predictive model. So in the resampled training results, the novel ensemble-based model learned different patterns—the test collection we had previously isolated to test the model output while splitting the entire dataset.

3.10 Performance Evaluation

The last step of the predictive analysis is to assess the novel model's output. This paper has measured novel models' performance with Precision, Kappa Static, Root Mean Square Error, Recall, F-measure, and accuracy. The explanation is that precision typically results in a false inference if the class distribution is unbalanced.

As we do not want a statistical model to skip the positive COVID-19 scenario, we want a recall score to be better than probable since this is a positive COVID-19 detective task. We should not ignore the accuracy since we can not predict the COVID-19 case even though it is positive.

4 Experiment Setup and Results

This novel ensemble-based classifier is implemented in python. Machine learning packages are utilized and installed in the Python IDE. Window 10 was placed on an i7 2.50 GHz computer with 16 GB RAM to analyze the effects of the novel ensemble-based classifier. The COVID-19 dataset is collected from the algorithm 1. The experiments are performed in two folds. (i) COVID-19 data is analyzed, and apply ensemble classifier to this data. The classifier's performance is compared with traditional classifiers, i.e. Decision Tree, ID3, and SVM, in Precision, Kappa Static, Root Mean Square Error, Recall, F-measure, and accuracy. (ii) The COVID-19 data is prepossessed, and size is reduced with class reconstruction, class decomposition, and standardization. The dataset is reduced to 10%, 20%, 30%, and 40%. The ensemble classifier is applied in various sizes of datasets.

There is a trade-off between classifier accuracy and its execution time when dataset size is varying. Classifier performs well concerning the accuracy and when the size of data is vast. On the other hand, it takes more execution time using a large dataset.

The COVID-19 dataset is further divided. The 66% of the data is used to train the classifier, while the rest of the remaining data checks its robustness. The proposed classifier is evaluated in both folds of experiments and measure the execution time of the classifier.

4.1 Precision

Precision is the percentage of the actual positive to all positive values. This is the assessment of patients and is accurately established among patients with a COVID-19 positive condition. Accuracy often offers a calculation of the related data. We must not begin managing a patient who may have COVID-19 negative. Figure 10 shows that the precision of the ensemble classifier is better than that
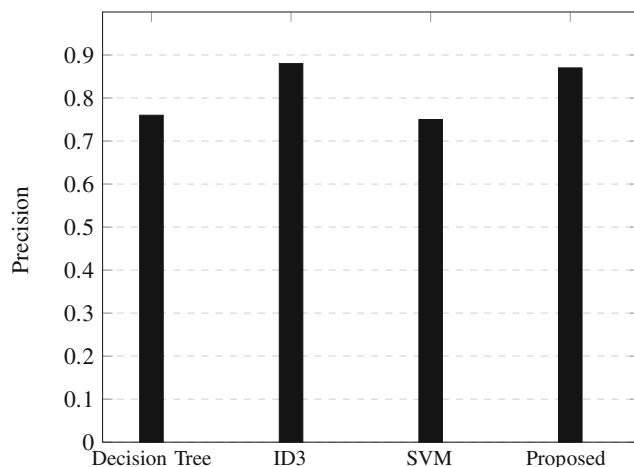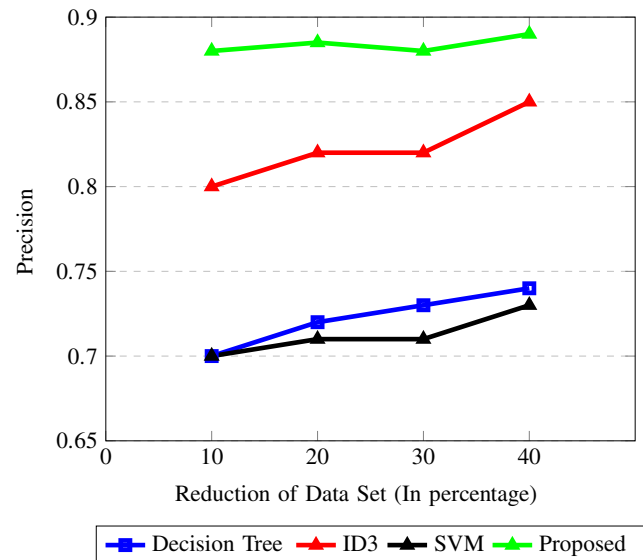

Fig. 11 Precision of classifiers at different dataset size

of other classifiers. Figure 11 shows the proposed novel classifier's precision when dataset size is reduced. Again the novel classifier performs well as compare to traditional classifiers.

4.2 Kappa Static

The kappa data is a calculation of how well machine learning categorised instances matched the data labelled as simple reality and tracked the accuracy of a random classification as calculated by the predicted accuracy. In addition to illuminating the way the novel classifier handled itself, the kappa statistics for one model is closely comparable with the kappa statistics for every other model that is used for the same classification assignment. As seen in the Fig. 12, the proposed novel classifier is stronger than other conventional
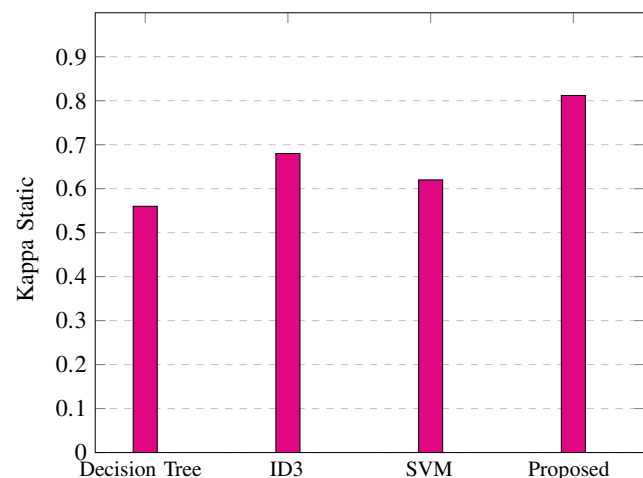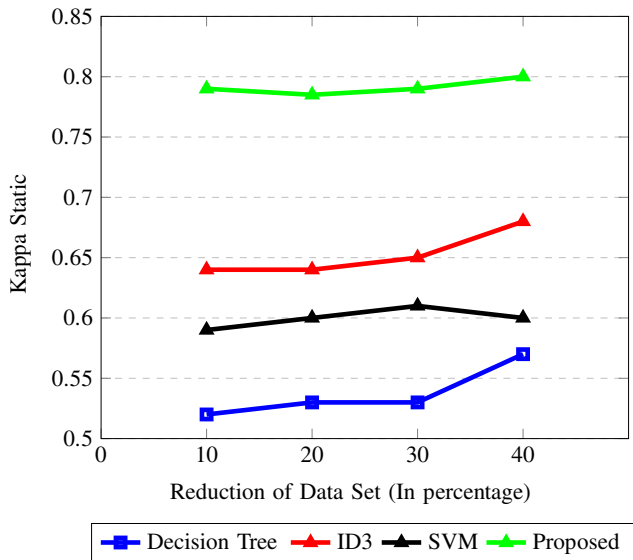

Fig. 10 Precision of classifiers

Fig. 12 Kappa Static of classifiers

Fig. 13 Kappa Static of classifiers at different dataset size

Fig. 15 Root Mean Square Error of classifiers at different dataset size

classifiers. Figure 13 shows the kappa static of proposed novel classifier in different size of data set. The proposed novel classifier does better than every other classifier.

4.3 Root Mean Square Error

Root Mean Square Error is measured as the square root of the mean of observed effects and projections. The squaring of each error causes positive values and returns the error metric to the initial units, contrasting with the square root of the mean squared error. Figure 14 shows that the proposed novel classifier has less Root Mean Square Error than the traditional classifier. Figure 15 shows the root mean square error of the proposed novel ensemble classifier in
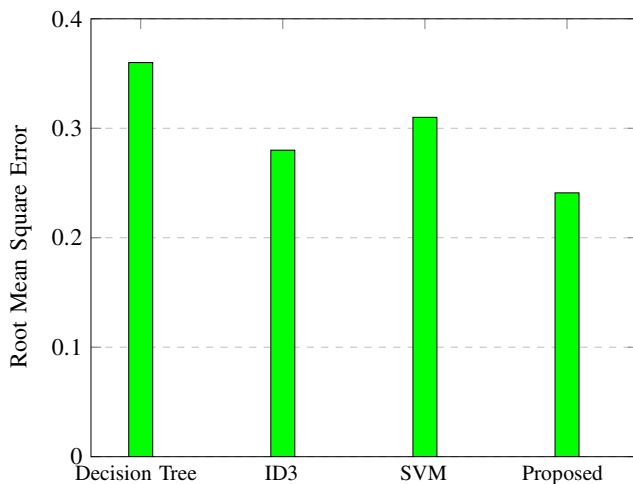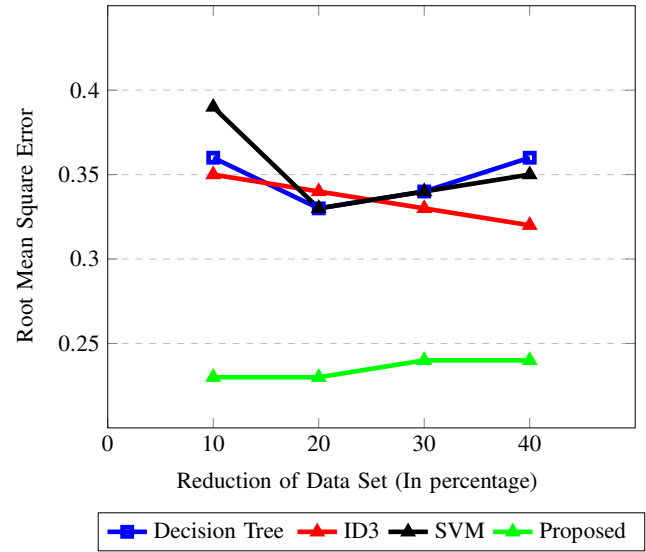
the reduced data set. The errors are always less in our novel classifier with a comparison of traditional classifiers.

4.4 Recall

The reminder is the calculation of the excellent qualities of our paradigm. So, note how many of us are accurately classified as getting a patient condition in all people with COVID-19 disease. Recall also demonstrates how effective our model will classify the related details. The novel ensemble-based classifier reaches a fair value of recall, as shown in Fig. 16, which is better than the other classifiers. On the other hand, it always performs better with reduced data set, as shown in Fig. 17.
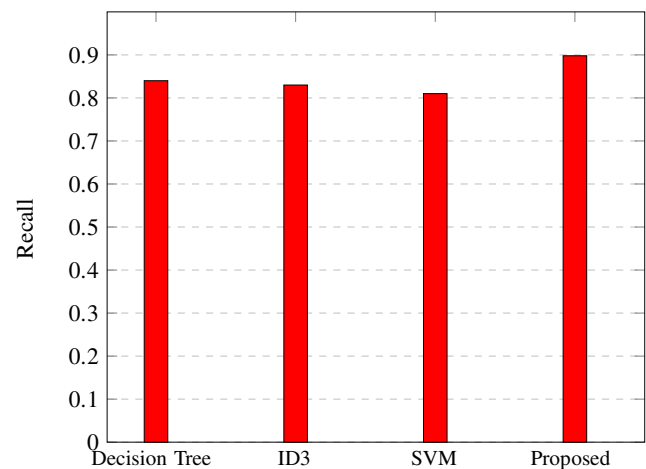

Fig. 14 Root Mean Square Error

Fig. 16 Recall of classifiers

Fig. 17 Recall of classifiers at different dataset size

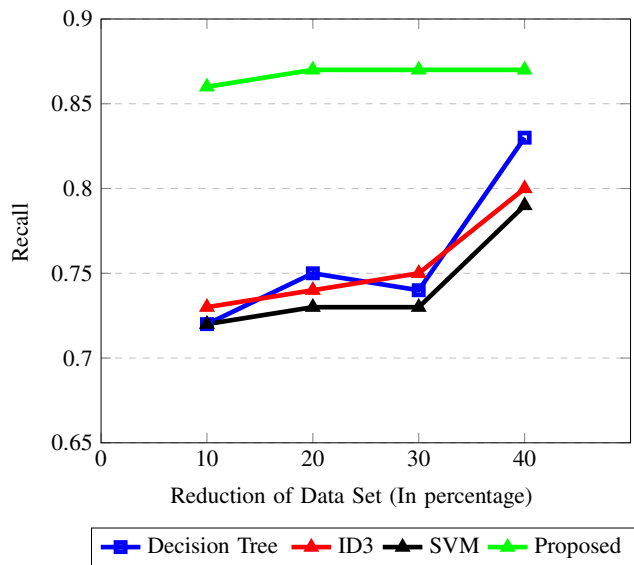Fig. 19 F-Measure of classifiers at different dataset size

4.5 F-measure

This is the estimate of a model's impact on a dataset. It may be used to assess classification schemes that designate examples as either "positive" or "negative." The F-score incorporates the F-precision measure's and recall and is a harmonic mean of the F-precision measures and recall. Figure 18 shows the F-measure of the proposed novel classifier, which is superior to other classification methods. Figure 19 shows the F-measure of the proposed novel classifier with different data sets, and it always performs better.

4.6 Accuracy

This is what we generally refer to as precision. P is the percentage of the observations that are accurate. In this way,
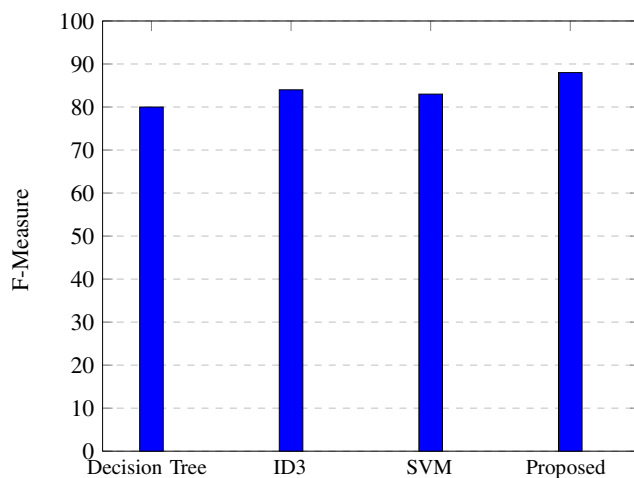
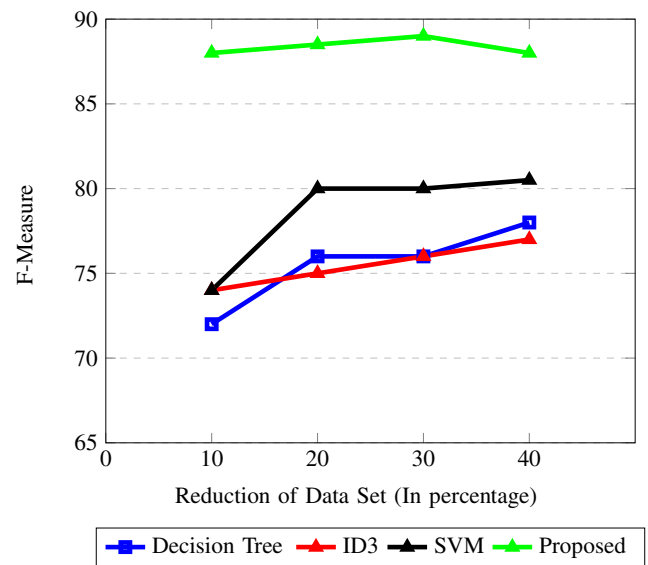a binomial test will be true only if the number of successes for both of the groups is the same. Figure 20 shows the accuracy of proposed novel ensemble classifier which is always better with conventional algorithms. Figure 21 depicts the accuracy of proposed novel classifier with reduced data set. The figure clearly shows the betterment of proposed novel classifier.

4.7 Execution Time

The execution time of a classifier is the period it takes to be executed. It is directly proportional to the size of the input data. The execution time is recorded only for the proposed novel classifier with different sizes of data set. Figure 22 shows the execution time of classifier with varying size.
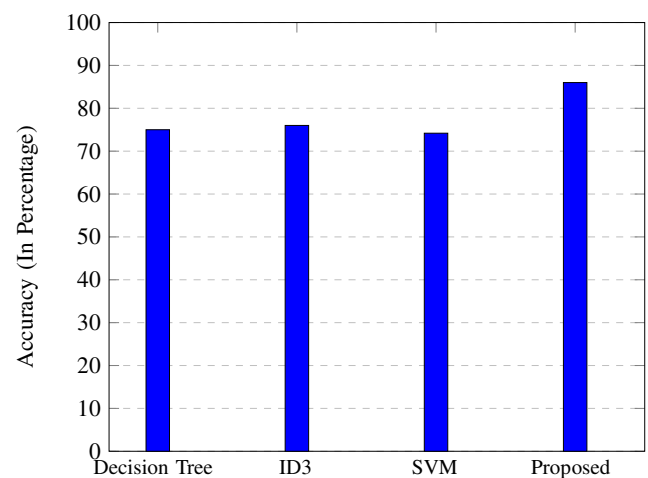

Fig. 18 F-Measure of classifiers
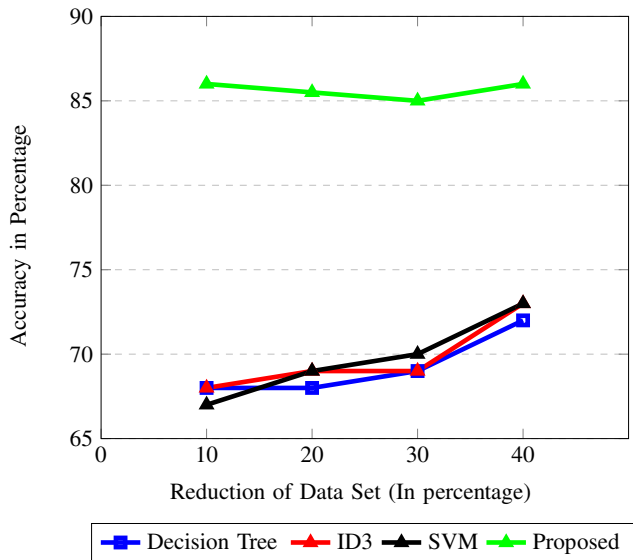

Fig. 20 Classifiers Accuracy

Fig. 21 Classifiers Accuracy at different dataset size

The Figure clearly shows that it is directly proportional to the size of the input data. When the data set is reduced, the execution time is also reduced.

5 Discussion

For the healthcare sector, valuable studies are identified. Firstly, very costly tests and medical instruments are avoided. Many simulations can be run, and the results can still be measured or queried. Secondly, it will explain the health status of COVID-19 patients for patients. Similarly, their care will allow physicians and researchers to know better the vulnerable points to launch their therapy. Third, machine learning in healthcare research is easier for the
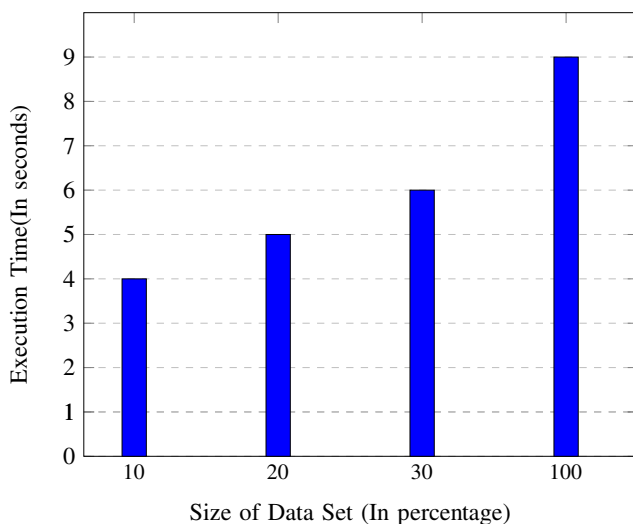

Fig. 22 Execution Time

general population to understand. The public should now be more educated about health literacy. Machine learning can also be used to identify COVID-19 in the field of education and training. It will also save lecturers time to describe malignant tumours repeatedly. The launch of a modern ensemble-based health and science classifier is typically intended to help more people understand complex issues.

6 Conclusion

The coronavirus epidemic had affected more than 4.4 million citizens in more than 200 countries and territories when this paper was written. In this paper, we applied machine learning techniques to predict whether a citizen is COVID-19 positive or not. A novel cost-sensitive ensemble-based classifier is applied for future work by taking classification costs into account. Besides, two approaches are introduced: one to improve the accuracy of the proposed novel ensemble-based classifier and improve the proposed novel classifier's speed by reducing the size of the dataset. The Class Reconstruction approach creates new class labels by calculating convex hulls of each class's clusters and using a convex hull-based algorithm to identify which clusters can be merged before assigning class labels. The goal of this algorithm is to improve the accuracy of the classifier. Accuracy can be improved by tweaking the gamma variable that allows for minimal geometric interference; however, no consistent method could quickly identify the best value for gamma, short of running the convex hull-based algorithm on multiple different gamma values.

Current limitations in applying COVID-19 analysis in clinical practice include a lack of better understanding of the relationships between COVID-19 biomarkers and pathological conditions, lack of standardization and normalization of procedures for the sampling, preparation, and analysis of COVID-19 data samples. In this type of study design, it is not easy to obtain a large enough representative sample of the overall population. It usually represents only a group of the total population.

In this paper, the symptom-based dataset of COVID-19 is processed. In future, deep learning can be applied to image data of COVID-19 for analysis. A generalized ensemble-based classifier may be built to address different forms of diseases such as swine flu, AIDs, cancer, etc.

References

Abdel-Basset, M., Chang, V., Hawash, H., Chakrabortty, R.K., Ryan, M. (2020a). Fss-2019-ncov: A deep learning architecture for semi-supervised few-shot segmentation of covid-19 infection. *Knowledge-Based Systems*, 106647.

Abdel-Basset, M., Chang, V., Mohamed, R. (2020b). Hsma_woa: A hybrid novel slime mould algorithm with whale optimization algorithm for tackling the image segmentation problem of chest x-ray images. *Applied Soft Computing*, *95*, 106642.

Abdel-Basset, M., Chang, V., Nabeeh, N.A. (2020c). An intelligent framework using disruptive technologies for covid-19 analysis. *Technological Forecasting and Social Change*, 120431.

Abdel-Basset, M., Chang, V., Hawash, H., Chakrabortty, R.K., Ryan, M. (2021). Fss-2019-ncov: A deep learning architecture for semi-supervised few-shot segmentation of covid-19 infection. *Knowledge-Based Systems*, *212*, 106647.

Abdel-Basset, M., Chang, V., Nabeeh, N.A. (2021). An intelligent framework using disruptive technologies for covid-19 analysis. *Technological Forecasting and Social Change*, *163*, 120431.

Ahn, N.-Y., Park, J.E., Lee, D.H., Hong, P.C. (2020). Balancing personal privacy and public safety in covid-19: Case of korea and france. arXiv:2004.14495.

Angurala, M., Bala, M., Bamber, S.S., Kaur, R., Singh, P. (2020). An internet of things assisted drone based approach to reduce rapid spread of covid-19. *Journal of Safety Science and Resilience*, *1*(1), 31–35.

Baudier, P., Kondrateva, G., Ammi, C., Chang, V., Schiavone, F. (2021). Patients perceptions of teleconsultation during covid-19: A cross-national study. *Technological Forecasting and Social Change*, *163*, 120510.

Benreguia, B., Moumen, H., Merzoug, M.A. (2020). Tracking covid-19 by tracking infectious trajectories. arXiv:2005.05523.

Cao, R., Tang, Z., Liu, C., Veeravalli, B. (2019). A scalable multicloud storage architecture for cloud-supported medical internet of things. *IEEE Internet of Things Journal*, *7*(3), 1641–1654.

Chamola, V., Hassija, V., Gupta, V., Guizani, M. (2020). A comprehensive review of the covid-19 pandemic and the role of iot, drones, ai, blockchain, and 5g in managing its impact. *IEEE Access*, *8*, 90225–90265.

Chang, V. (2018a). Computational intelligence for medical imaging simulations. *Journal of Medical Systems*, *42*(1), 10.

Chang, V. (2018b). Data analytics and visualization for inspecting cancers and genes. *Multimedia Tools and Applications*, *77*(14), 17693–17707.

Chang, V. (2018c). An overview, examples, and impacts offered by emerging services and analytics in cloud computing virtual reality. *Neural Computing and Applications*, *29*(5), 1243–1256.

Chen, S., Yang, J., Yang, W., Wang, C., Bärnighausen, T. (2020). Covid-19 control in china during mass population movements at new year. *The Lancet*, *395*(10226), 764–766.

De Santis, E., Martino, A., Rizzi, A. (2020). An infoveillance system for detecting and tracking relevant topics from italian tweets during the covid-19 event. *IEEE Access*, *8*, 132527–132538.

Dhiman, G. (2019a). ESA: A Hybrid Bio-inspired Metaheuristic Optimization Approach for Engineering Problems. *Engineering with Computers*, 1–31.

Dhiman, G. (2019b). MOSHEPO: A Hybrid Multi-objective Approach to Solve Economic Load Dispatch and Micro Grid Problems. *Applied Intelligence*, 1–19.

Dhiman, G., & Kaur, A. (2019c). A Hybrid Algorithm based on Particle Swarm and Spotted Hyena Optimizer for Global Optimization. In *Soft Computing for Problem Solving* (pp. 599–615): Springer.

Dhiman, G., & Kaur, A. (2019d). STOA: A Bio-inspired based Optimization Algorithm for Industrial Engineering Problems. *Engineering Applications of Artificial Intelligence*, *82*, 148–174.

Dhiman, G., & Kaur, A. (2020a). HKn-RVEA: A Novel Many-objective Evolutionary Algorithm for Car Side Impact Bar Crash-worthiness Problem. *International Journal of Vehicle Design*.

Dhiman, G., Soni, M., Pandey, H.M., Slowik, A., Kaur, H. (2020b). A novel hybrid hypervolume indicator and reference vector adaptation strategies based evolutionary algorithm for many-objective optimization. *Engineering with Computers*, 1–19.

Dhiman, G., & Garg, M. (2020c). MoSSE: A Novel Hybrid Multi-objective Meta-heuristic Algorithm for Engineering Design Problems. *Soft Computing*.

Dhiman, G., Chang, V., Kant Singh, K., Shankar, A. (2021). Adopt: automatic deep learning and optimization-based approach for detection of novel coronavirus covid-19 disease using x-ray images. *Journal of Biomolecular Structure and Dynamics*, 1–13.

Elhadad, M.K., Li, K., Gebali, F. (2020). Detecting misleading information on covid-19. *Ieee Access*, *8*, 165201–165215.

Farahani, B., Firouzi, F., Chang, V., Badaroglu, M., Constant, N., Mankodiya, K. (2018). Towards fog-driven iot ehealth: Promises and challenges of iot in medicine and healthcare. *Future Generation Computer Systems*, *78*, 659–676.

Garg, M., & Dhiman, G. (2020). A Novel Content Based Image Retrieval Approach for Classification using GLCM Features and Texture Fused LBP Variants. *Neural Computing & Applications*.

Gomathi, S., Kohli, R., Soni, M., Dhiman, G., Nair, R. (2020). Pattern analysis: predicting covid-19 pandemic in india using automl. *World Journal of Engineering*.

Gupta, V., Singh Gill, H., Singh, P., Kaur, R. (2018). An energy efficient fog-cloud based architecture for healthcare. *Journal of Statistics and Management Systems*, *21*(4), 529–537.

He, X., Chu, L., Qiu, R.C., Ai, Q., Ling, Z. (2018). A novel data-driven situation awareness approach for future gridsusing large random matrices for big data modeling. *IEEE Access*, *6*, 13855–13865.

Hussain, A.A., Bouachir, O., Al-Turjman, F., Aloqaily, M. (2020). Ai techniques for covid-19. *IEEE Access*, *8*, 128776–128795.

Imran, A.S., Daudpota, S.M., Kastrati, Z., Batra, R. (2020). Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on covid-19 related tweets. *IEEE Access*, *8*, 181074–181090.

Jamshidi, M., Lalbakhsh, A., Talla, J., Peroutka, Z., Hadjilooei, F., Lalbakhsh, P., Jamshidi, M., La Spada, L., Mirmozafari, M., Dehghani, M., et al. (2020). Artificial intelligence and covid-19: deep learning approaches for diagnosis and treatment. *IEEE Access*, *8*, 109581–109595.

Kaur, S., Awasthi, L.K., Sangal, A.L., Dhiman, G. (2020). Tunicate swarm algorithm: A new bio-inspired based metaheuristic paradigm for global optimization. *Engineering Applications of Artificial Intelligence*, *90*, 103541.

Kaur, M., Kaur, R., Singh, N., Dhiman, G. (2021). Schoa: a newly fusion of sine and cosine with chimp optimization algorithm for hls of datapaths in digital filters and engineering applications. *Engineering with Computers*, 1–29.

Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., Hoffmann, M. (2014). Industry 4.0. *Business & Information Systems Engineering*, *6*(4), 239–242.

Latif, S., Usman, M., Manzoor, S., Iqbal, W., Qadir, J., Tyson, G., Castro, I., Razi, A., Boulos, M.N.K., Weller, A., et al. (2020). Leveraging data science to combat covid-19: A comprehensive review.

Li, K. (2020). Hierarchical pooling strategy optimization for accelerating asymptomatic covid-19 screening. *IEEE Open Journal of the Computer Society*, *1*, 276–284.

Li, Y., Wei, D., Chen, J., Cao, S., Zhou, H., Zhu, Y., Wu, J., Lan, L., Sun, W., Qian, T., et al. (2020). Efficient and effective training of covid-19 classification networks with self-supervised dual-track learning to rank. *IEEE Journal of Biomedical and Health Informatics*, *24*(10), 2787–2797.

Liu, N., & Wang, H. (2010). Ensemble based extreme learning machine. *IEEE Signal Processing Letters*, *17*(8), 754–757.

Marmarelis, V.Z. (2020). Predictive modeling of covid-19 data in the us: Adaptive phase-space approach. *IEEE Open Journal of Engineering in Medicine and Biology*, *1*, 207–213.

Nair, R., Gupta, S., Soni, M., Shukla, P.K., Dhiman, G. (2020). An approach to minimize the energy consumption during blockchain transaction. *Materials Today: Proceedings*.

Niu, R., Wong, E.W.M., Chan, Y.-C., Van Wyk, M.A., Chen, G. (2020). Modeling the covid-19 pandemic using an seihr model with human migration. *IEEE Access*, *8*, 195503–195514.

Ouyang, X., Huo, J., Xia, L., Shan, F., Liu, J., Mo, Z., Yan, F., Ding, Z., Yang, Q., Song, B., et al. (2020). Dual-sampling attention network for diagnosis of covid-19 from community acquired pneumonia. *IEEE Transactions on Medical Imaging*.

Pham, Q.-V., Nguyen, D.C., Hwang, W.-J., Pathirana, P.N., et al. (2020). Artificial intelligence (ai) and big data for coronavirus (covid-19) pandemic: A survey on the state-of-the-arts.

Quatieri, T.F., Talkar, T., Palmer, J.S. (2020). A framework for biomarkers of covid-19 based on coordination of speech-production subsystems. *IEEE Open Journal of Engineering in Medicine and Biology*, *1*, 203–206.

Qureshi, F.F., Iqbal, R., Qasim, M., Doctor, F., Chang, V. (2017). Integration of omni channels and machine learning with smart technologies. *Journal of Ambient Intelligence and Humanized Computing*, 1–17.

Ren, J., Yan, Y., Zhao, H., Ma, P., Zabalza, J., Hussain, Z., Luo, S., Dai, Q., Zhao, S., Sheikh, A., et al. (2020). A novel intelligent computational approach to model epidemiological trends and assess the impact of non-pharmacological interventions for covid-19. *IEEE Journal Of Biomedical and Health Informatics*, *24*(12), 3551–3563.

Sharma, A., Zanotti, P., Musunur, L.P. (2020). Drive through robotics: Robotic automation for last mile distribution of food and essentials during pandemics. *IEEE Access*, *8*, 127190–127219.

Singh, K.D., & Sood, S.K. (2020a). 5g ready optical fog-assisted cyber-physical system for iot applications. *IET Cyber-Physical Systems: Theory & Applications*, *5*(2), 137–144.

Singh, P., & Kaur, R. (2020b). An integrated fog and artificial intelligence smart health framework to predict and prevent covid-19. *Global Transitions*, *2*, 283–292.

Singh, R.P., Javaid, M., Haleem, A., Vaishya, R., Al, S. (2020c). Internet of medical things (iomt) for orthopaedic in covid-19 pandemic: Roles, challenges, and applications. *Journal of Clinical Orthopaedics and Trauma*.

Small, M., & Cavanagh, D. (2020). Modelling strong control measures for epidemic propagation with networks–a covid-19 case study. arXiv:2004.10396.

Sood, S.K., & Singh, K.D. (2018). An optical-fog assisted eeg-based virtual reality framework for enhancing e-learning through educational games. *Computer Applications in Engineering Education*, *26*(5), 1565–1576.

Ulhaq, A., Born, J., Khan, A., Gomes, D.P.S., Chakraborty, S., Paul, M. (2020). Covid-19 control by computer vision approaches: A survey. *IEEE Access*, *8*, 179437–179456.

Wang, L., Ma, Y., Yan, J., Chang, V., Zomaya, A.Y. (2018). pipscloud: High performance cloud computing for remote sensing big data management and processing. *Future Generation Computer Systems*, *78*, 353–368.

Wang, J., Bao, Y., Wen, Y., Lu, H., Luo, H., Xiang, Y., Li, X., Liu, C., Qian, D. (2020a). Prior-attention residual learning for more discriminative covid-19 screening in ct images. *IEEE Transactions on Medical Imaging*.

Wang, Z., Liu, Q., Dou, Q. (2020b). Contrastive cross-site learning with redesigned net for covid-19 ct classification. *IEEE Journal of Biomedical and Health Informatics*, *24*(10), 2806–2813.

Xu, X., Mo, R., Yin, X., Khosravi, M.R., Aghaei, F., Chang, V., Li, G. (2020). Pdm: Privacy-aware deployment of machine-learning applications for industrial cyber-physical cloud systems. *IEEE Transactions on Industrial Informatics*.

Yang, Y., Zheng, X., Guo, W., Liu, X., Chang, V. (2019). Privacy-preserving smart iot-based healthcare big data storage and self-adaptive access control system. *Information Sciences*, *479*, 567–592.

You, Z., & Feng, L. (2020). Integration of industry 4.0 related technologies in construction industry: a framework of cyber-physical system. *IEEE Access*, *8*, 122908–122922.

Prabh Deep Singh presently doing a Ph.D. from Punjabi University, Patiala. He has completed his M.E. from Thapar University, Patiala. He has published more than 25 papers in International Journals. And more than 20 papers in National / International Conferences. His work is published and cited in highly reputed journals. His area of research is smart city development, Big Data, IoT, Artificial Intelligence, Cloud computing, Fog computing.

Rajbir Kaur has completed her Ph.D. in Signal System. She is currently working as Assistant Professor, Electronics and Communication Engineering, Punjabi University, Patiala. She has published more than 40 papers in National/International Journals and Conferences. Her work is published and cited in highly reputed journals like Elsevier, Springer, etc. Her areas of research are Signal systems, Cloud computing, Fog computing, and Edge computing.

Kiran Deep Singh has completed his Doctorate in Computer Science and Engineering from Guru Nanak Dev University, Amritsar. He received his M.E. degree in Software Engineering from Thapar University, Patiala. He is currently working as Head, Department of Computer Science and Engineering, Khalsa College of Engineering and Technology, Amritsar. He has published many research papers in highly reputed journals. His current working research areas include the Internet of Things (IoT), Big data, and Fog/cloud computing.

Gaurav Dhiman received his Ph.D. in Computer Engineering from Thapar Institute of Engineering & Technology, Patiala. He has completed his Master Degree of Computer Applications from Thapar Institute of Engineering & Technology, Patiala. He is currently working as an Assistant Professor with the Department of Computer Science, Government Bikram College of Commerce, Patiala. He was selected as an outstanding reviewer from Knowledge-Based Systems (Elsevier). He has published more than 100 peer-review research papers (indexed in SCI-SCIE) and 10 international books. He is the member of IEEE. He is also serving as the guest editor of more than ten special issues in various peer-reviewed journals (including IEEE Transactions on Industrial Informatics). He is also in editorial board members of Current Chinese Computer Science (Bentham Science), Current Dissertations (Bentham Science), Journal of Fuzzy Logic and Modeling in Engineering (Bentham Science), Journal of the Institute of Electronics and Computer (IEC Science), Journal of PeerScientist, Artificial Intelligence Evolution, Acta Scientific Computer Sciences. His research interest includes Single-, Multi-, Many-objective optimization (Bio-inspired, Evolutionary and Quantum), Soft computing (Type-1 and Type-2 fuzzy sets), Power systems, and Change detection using remotely sensed high-resolution satellite data. His research articles can be found in Knowledge-Based Systems (Elsevier), Advances in Engineering Software (Elsevier), Applied Intelligence (Springer), Journal of Computational Science (Elsevier), Engineering with Computers (Springer), Applied Soft Computing (Elsevier), Engineering Applications of Artificial Intelligence (Elsevier), Journal of Ambient Intelligence and Humanized Computing (Springer), among others. His research can also be seen in http://www.dhimangaurav.com.