

Access this article online
Quick Response Code:

Website: www.jehp.net
DOI: 10.4103/jehp.jehp_298_23

Analysis of computational intelligence approaches for predicting disease severity in humans: Challenges and research guidelines

Geetha Narasimhan, Akila Victor

Abstract:

The word disease is a common word and there are many diseases like heart disease, diabetes, breast cancer, COVID-19, and kidney disease that threaten humans. Data-mining methods are proving to be increasingly beneficial in the present day, especially in the field of medical applications; through the use of machine-learning methods, that are used to extract valuable information from healthcare data, which can then be used to predict and treat diseases early, reducing the risk of human life. Machine-learning techniques are useful especially in the field of health care in extracting information from healthcare data. These data are very much helpful in predicting the disease early and treating the patients to reduce the risk of human life. For classification and decision-making, data mining is very much suitable. In this paper, a comprehensive study on several diseases and diverse machine-learning approaches that are functional to predict those diseases and also the different datasets used in prediction and making decisions are discussed in detail. The drawbacks of the models from various research papers have been observed and reveal countless computational intelligence approaches. Naïve Bayes, logistic regression (LR), SVM, and random forest are able to produce the best accuracy. With further optimization algorithms like genetic algorithm, particle swarm optimization, and ant colony optimization combined with machine learning, better performance can be achieved in terms of accuracy, specificity, precision, recall, and specificity.

Keywords:

Accuracy, ant colony optimization, computational intelligence, genetic algorithm, machine learning, medical application, particle swarm optimization, precision, prediction, recall, specificity

Introduction

Many different diseases caused infections and affected many people's lives and made it very difficult for living. It took many years to come across those critical situations. Diseases that have been discovered across a period can be stated as pandemic and epidemic diseases.^[1] If the disease and its illness persist for over a period that is not normal in a particular area, then it can be said as an epidemic.^[2] The infectious disease outbreak can be said as pandemic which is

the reason for the increase in infection and death in that location.

The growth of diseases leads to an increase in medical data. It is very hard to process and investigate those vast volumes of data. There are also many issues such as privacy and security, analyzing the data effectively in treating the medical data. Advanced techniques such as artificial intelligence (AI) help the researcher solve difficult problems. The machine is trained in a way to think and act like a human. These AI also help in solving complex data problems. AI solves all kinds of big data

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Narasimhan G, Victor A. Analysis of computational intelligence approaches for predicting disease severity in humans: Challenges and research guidelines. J Edu Health Promot 2023;12:334.

School of Computer
Science and Engineering,
Vellore Institute of
Technology, Vellore,
Tamil Nadu, India

Address for correspondence:

Dr. Akila Victor,
School of Computer
Science and Engineering,
Vellore Institute of
Technology, Vellore,
Tamil Nadu, India.
E-mail: akilavictor@vit.
ac.in

Received: 02-03-2023
Accepted: 12-04-2023
Published: 29-09-2023

problems by making the decision easier for complex data. These AI can be used in the field of health care in solving complex problems. Machine learning is one such field of AI that is used in the field of medicine to predict and diagnose various diseases in advance to treat the patients prior and treat them on time to save from severity.^[3]

The machine-learning approaches such as K nearest neighbor (KNN), support vector machine (SVM), random forest (RF), decision tree (DT), logistic regression (LR), Linear discriminant analysis (LDA), multilayer perceptron (MLP), principal component analysis (PCA), ensemble voting algorithm, genetic algorithm (GA), particle swarm optimization (PSO), dragonfly algorithm (DA), and artificial bee colony (ABC) were used in predicting and diagnosing the different diseases. These techniques were used by various authors as a single algorithm or combination of two or more in their model to predict the diseases.

The major focus of this work is to have a detailed study of various papers that used machine-learning techniques to solve and predict the disease; also, to analyze the different evaluation parameters used in these models and how it proves the model to be better. The work focuses on a different dataset for diseases like heart disease, diabetes, COVID-19, vitamin D deficiency, breast cancer, and kidney disease. The article is organized into different sections namely Materials and methods section which gives paper list and publications, various datasets used for predicting diseases using machine learning, evaluation criteria used for highlighting the accuracy of the model, study of different machine-learning approaches used in various papers, the different optimization techniques used for predicting the disease, and finally conclusion of entire work.

Materials and Methods

In this review paper, 91 articles have been included based on the diseases and the prediction methods followed by the author of the paper in the medical field. All the papers are cited from diverse publications such as IEEE, science direct, and Elsevier, which is shown in Figure 1.

Various classification and regression algorithms used in predicting different diseases have been studied and explained in this article. Table 1 gives the summary of the literature study that dealt with various machine-learning models used in the classification and prediction of different diseases like heart disease, breast cancer, diabetes, and COVID-19.

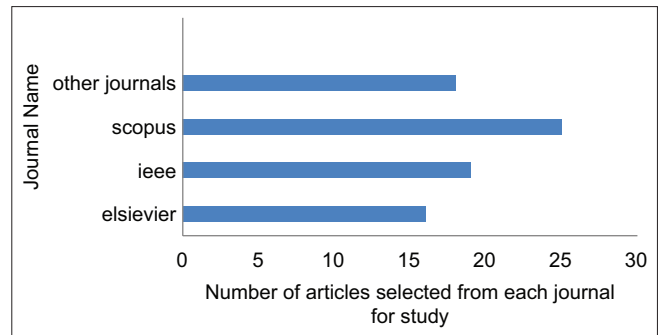


Figure 1: Journals cited in the article

The Various Datasets Used in the Medical Field for Computational Approaches

To perform classification and regression, it is important to have a dataset related to the diseases. The dataset is divided into the training and testing phases. Based on the training and testing, the performance of the algorithm is measured based on the evaluation metrics like accuracy, specificity, precision, recall, f score, and sensitivity. Based on the evaluation metrics, accuracy will be considered the suitable algorithm for that dataset and model. Researchers have used different kinds of datasets for each disease. Wisconsin dataset includes breast cancer disease data with nine features, Pima Indian diabetes dataset includes eight features, and one outcome variable. The heart disease dataset contains 14 features and 1 target feature. The liver disease consists of 10 features and 1 target feature. Various kinds of data that are used by the researchers for predicting, classifying, and COVID-19 forecasting are clinical data, data from online, and biomedical data. John Hopkins University data set is the most utilized data set by many researchers for forecasting and COVID-19 production. Information like positive cases, patients recovered, and death are available in this data set. This information can be gathered from the country as well as state and also available on Kaggle. Issues like data imbalance are faced by many researchers while working with this data set. The difference in intervention and the density of population are also considered as a limitation that affects the prediction. Table 2 specifies the different datasets used for the classification and prediction of disease.

From online, huge textual data can be obtained that can be used for predicting COVID-19 spread and growth. It is very important to understand the spread of the virus at the local and national levels in predicting the virus. Factors like differentiating between the death that happened by virus and normal death; sometimes data are unavailable, gathering data from local government needed to be considered when predicting COVID-19 with the help of online data. The biomedical data include chest images, X-rays, and tomography which are

used to overcome the disadvantages of RT-PCR. These datasets are used by many researchers in predicting COVID-19 by analyzing those images. These images also have their disadvantages and difficulties in predicting the virus. These datasets are not useful in producing accurate results to differentiate between pneumonia and COVID-19.

Evaluation Criteria

The various evaluation criteria used by different authors in their papers for measuring the performance are accuracy, sensitivity, precision, recall, *f* score, specificity, and area under the curve. Based on these evaluation metrics, the prediction is carried out. Table 3 lists the different evaluation criteria used in measuring the performance of an algorithm.^[50] Among the evaluation criteria, accuracy plays a major role in deciding the performance of the model. It refers to the ratio of the overall count of correct classified samples to the total number of correctly and wrongly classified.

The nearest measurement is referred to by precision, the ratio of true positives, and total predicted positives. Recall and sensitivity are similar. *f* Score denotes the accuracy of the test. The graph that shows the performance of the model used for classification is denoted by the receiver operating characteristic curve (ROC). The area under the curve (AUC) is the area covered by various thresholds that are used to calculate the labels of class. The model is said to be the best classification technique when higher values are achieved in inaccuracy, recall, precision, *f*-score, and AUC.

Study of Various Models Applied in Different Diseases

Process flow from the papers

The papers that are used in this review articles follow a process flow for the prediction of disease using different algorithms^[50] shown in Figure 2. Basically, the method starts with gathering inputs for performing future importance and preprocessing the dataset due to the missing and noisy data. The dataset is split as training and testing data. The training data is used by the model for identifying the pattern and evaluating the performance metrics mentioned in Table 3. The final output from the model is to predict the disease from the raw medical dataset.

Dahiwade *et al.*^[4] tried to predict the symptoms of patients using KNN algorithm and CNN. The experiment produced 84.5% accuracy for CNN which is more than KNN. In this, the author used a patient disease dataset for finding the symptoms. The execution time for KNN and CNN is compared and found CNN takes less time for computation and achieved 84.5% accuracy. The

author has not used any specific feature selection of the attributes. In the same way, Grover *et al.*^[5] tried to predict the disease based on the symptoms which helps the doctor diagnose the disease of the patient. The author has focused on two research questions. One is whether machine learning is suitable for predicting disease and the other one is to find the best-suited algorithm for prediction. Grover and his team answered those research questions and found that the convolutional neural network for the multimodal disease prediction algorithm achieved an accuracy of 94.8%. The author did not focus on any feature selection or extraction. Also, there is no comparison of algorithms handled.

Kohli and Arora^[6] applied the classification algorithm on three disease datasets, namely, heart, diabetes, and breast cancer. The authors used feature selection by backward *P*-value test. If the *P*-value > 0.05, then the variable is removed. Based on the feature selected, the classification algorithms such as LR, DT, RF, SVM, and adaptive boosting are applied. From the experiment, the authors found that LR produced 87.1% accuracy for heart disease, SVM with 85.71% for diabetes, and 98.57% for breast cancer from the AdaBoost classifier. Likewise, Kumar and Pathak^[7] applied DT, RF, Naïve Bayes, and KNN machine-learning algorithms for predicting the disease to help the patients. They found that Naïve Bayes achieved 95.21% accuracy. The authors developed a GUI which will store the data entered by the patients and the system can be used by all. The author did not focus on association rules and discretization techniques. There are many classification algorithms that best suit the model and they are not utilized in this study.

Saied *et al.*^[8] focused on LR, LDA, KNN, CART, GNB, and SVM algorithms for classification of Alzheimer's disease. The input is obtained with the help of radio frequency signals and machine learning for classification. The authors found that LR achieved 98.97% accuracy. The major drawback is that the accuracy achieved with minimal observations is due to the limited voxel model in CST. Kumar *et al.*^[9] focused on COVID-19, heart, and diabetes disease. Various machine-learning algorithms are applied and found that the proposed model performed better compared to other machine-learning algorithms. A mobile application was developed and the disease symptoms are given based on that the prediction is done. The proposed model reached 1.48% of the measure. With the proposed model, the author focused on only disease prediction and did not concentrate on metaheuristic techniques for tuning the parameters. In the same way, kidney disease is also predicted using DTs with a dataset collected from the UCI repository which has 24 and 1 class attributes. The DT helps the dataset to get divided into subpopulations. A later regression tree is applied to give data to the branches,

Table 1: Summary of various papers that used medical data for the prediction of disease

Reference paper	Method used/model applied	Result
[4]	KNN, CNN	CNN achieved 84.5% more than KNN
[5]	CNN MDRP, CNN UDRP	CNN MDRP 94.8%; more than CNN UDRP
[6]	LR, DT, RF, SVM, adaptive boosting	Logistic regression achieved 87.1% for heart disease detection, SVM with 85.71% for diabetes prediction, and AdaBoost with 98.57% for breast cancer detection
[7]	DT, RF, NB, KNN	NB: 95.21% DT: 95.12% RF: 95.11% KNN: 95.12%
[8]	Logistic regression, LDA, KNN, CART, GNB, SVM	Logistic regression: 98.97% LDA: 95.56% KNN: 58% CART: 78.667% GNB: 43.33% SVM: 21.33%
[9]	Machine-learning models	F_{measure} : 1.4765% for COVID dataset
[10]	Linear SVM, logistic regression, an ensemble method	Ensemble model performed better than LSVM, LR
[11]	KNN, DT, RF, bagging classifier, extra trees, SGD, gradient boosting, SVM, MLP	RF: 96%
[12]	SVM for predicting the model, logistic regression for classification, NB, DT	NB achieved higher accuracy, processing time
[13]	PCA, MLP	Ten-fold cross-validation: 86.7% accuracy
[14]	SGD classifier, KNN, RF, logistic regression, hard-voting ensemble method	Hard voting ensemble method: 90%
[15]	Logistic regression, SVM, KNN	Heart disease: KNN 84% Breast cancer: logistic regression 95% Diabetes: logistic regression 77%
[16]	Heart disease: NB, DT with PCA without PCA-SVM diabetes: WEKA tool	Heart disease: NB, DT with PCA without PCA-SVM diabetes: WEKA tool
[17]	SVM, ANN	Fused ML achieved 94.87% accuracy
[18]	CNN UDRP	CNN UDRP: 94.80% accuracy
[19]	AdaBoost, DT, logistic regression, RF, SVM	LDA: 80%
[20]	SVM, RF	The accuracy of 99.35%, 99.37%, and 99.14% on diabetes, kidney, and liver disease
[6]	AdaBoost, DT, RF, SVM, logistic regression	Heart disease: logistic regression: 87.1% Diabetes-SVM: 85.71% Breast cancer: Adaboost: 98.57%
[21]	SVM, DT	SVM: 85% DT: 83%
[22]	PCA, SVM, wavelet decomposition for preprocessing	Accuracy of 82.6% Error rate of 0.174 Sensitivity of 1.0 Specificity of 0.888
[23]	RF, NB, SVM, Hoeffding decision tree, logistic, model tree	Random forest: 95.08% Naïve Bayes: 93.44% SVM: 90.16% Hoeffding decision tree: 81.24% Logistic model tree: 80.69%
[24]	ANN, SVM, RF	Better performance is achieved by random forest
[25]	ML algorithms	RF performed better with 80% AUC
[26]	ML classification techniques	SVM performed better with 60% accuracy
[27]	ML-based improved model	The data-driven technique has higher accuracy
[28]	Comparison between ML and soft computing models	MLP and ANFIS have better accuracy
[29]	Random forest technique with AdaBoost algorithm	94% accuracy has been achieved by this model
[30]	LR, RF, XGB	Among the classification algorithm, RF performed well with 95% accuracy

Contd...

Table 1: Contd...

Reference paper	Method used/model applied	Result
[31]	ML models and ensemble learning techniques	The higher accuracy is obtained by ensemble learning techniques
[32]	Traditional ML model and federated learning models	SoftMax activation function and SGD optimizer give better accuracy
[33]	ML models	99.93% accuracy obtained by ARMA
[34]	ARIMA, SVR, NN, LR	To decrease the spread of COVID-19, preventive measures have to be followed
[35]	SVM, ANN, RF, DT, LR, KNN	The overall accuracy of 89.98% was achieved by the model
[36]	LR, DT, SVM, NB, ANN	DT has higher accuracy of 94.99%
[37]	Elman neural network, SVM	SVM along with Fuzzy granulation performs well in the prediction
[38]	ML	To result has high accuracy for the dataset taken
[39]	LR, MLP, vector autoregression	MLP provides better results compared to other methods
[40]	SVM	Compared to the neural network, SVM performed better with 98.88% accuracy
[41]	LR, MLP, RF, SVM	The performance of the algorithms is compared and found that SVM is effective
[42]	SVM, rough set, Bayesian ridge, polynomial regression, RNN, the SIR model	RNN was the most effective
[43]	SVM	The method is more effective and accurate
[44]	LR, SVM, PR, LSTM	The sigmoid model produces higher accuracy
[45]	LR, SVM, LASSO, ES	A better result was achieved by ES
[46]	RF, linear model, SVM, NN	RF performed better
[47]	Ensemble ML, BOW, TF/IDF, LR, MNB	MNB and LR showed better results
[48]	LR, PR, Holts method	The methods obtained a 0.99 R^2 value
[49]	Polynomial regression	PR predicts the new cases better

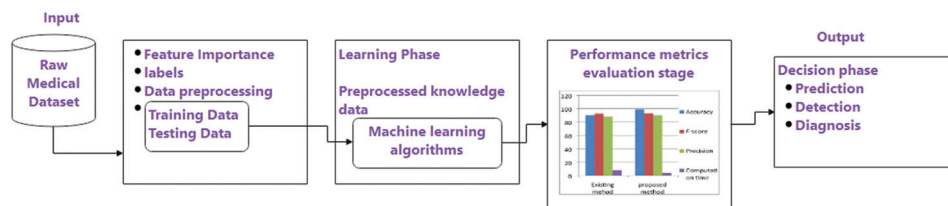


Figure 2: Flow of process from the papers^[3]

and based on squared deviation, the partition is done.^[52] In another paper, the authors^[53] used heart disease data for predicting the disease using J48 DT, NB, and Bagging techniques. Among the 13 features, 11 were selected for further investigation. The experimental result showed 85.03% for the bagging technique. The same authors^[54] used NB, J48, and RBF for breast cancer data and found 97.4% accuracy by Naïve Bayes. The authors used the WEKA tool for performance evaluation. The other evaluation metrics such as sensitivity and specificity were also considered for performance testing. The paper did not focus on developing an algorithm for the dataset taken. Also, feature selection techniques were not incorporated.

Syed *et al.*^[10] performed LR and SVM for predicting Alzheimer’s disease and found that the ensemble model performed better compared to linear regression and linear SVM. *T*-test analysis was done based on a 5% significance level. A web application was developed for

predicting the system. The authors have not implemented any kind of feature reduction techniques for reducing the attributes. Sambasivam *et al.*^[11] focused on predicting vitamin D deficiency. The machine-learning techniques such as KNN, RF, DT, AdaBoost, SGD, ET, BC, SVM, GB, and MLP were used and compared to find the best-suited algorithm for prediction. The experimental result shows that the RF classifier obtained 96% accuracy. The system uses *P* value of the variables using *t*-test. The author could have applied the model for other diseases to predict based on the symptoms. Also feature reduction is not applied. Ferjani^[55] worked on predicting different diseases using the DT, NB, KNN, SVM, LR, RF, and CNN. The author found that in predicting kidney and Parkinson disease, SVM performed better. For heart disease LR and breast cancer RF, CNN performed better. The author did not try any complex algorithms for prediction. Also feature selection or reduction techniques have not been applied.

Table 2: Different datasets used for the classification and prediction of disease

Disease	Dataset	Attributes	Total size
Heart disease	Coronary heart disease dataset	14 attributes and 1 target variable	1025 records
Breast cancer	Wisconsin breast cancer dataset	32 features	570 records
Liver disease	Liver disease dataset	10 features and 1 target feature	30,692 records
Coronary heart disease	Taken from UCI repository and standard dataset	76 attributes	303 records
Diabetes	PIMA	8 feature and 1 outcome variable	769 records
Vitamin D deficiency	3044 college students	11 parameters	–
COVID-19	John Hopkins university	7 features	306,429 records
Kidney disease	UCI repository	24 feature and 1 class label	400 instances

Table 3: Evaluation criteria used for measuring the performance of computational intelligence approaches^[51]

Measure	Formula	Description
Precision	$\frac{TP}{TP + FP}$	The number of positives to that of total positives predicted
Recall	$\frac{TP}{TP + FN}$	Number of positives predicted from total positives from the sample
Specificity	$\frac{TN}{TN + FP}$	Number of results negative from the total number of samples with no disease
Accuracy	$\frac{TP + TN}{TP + FP + FN + TN}$	Number of predictions that are correct from the entire predictions
f-Score	$2 * \left(\frac{Precision * Recall}{Precision + Recall} \right)$	This criterion is used to solve problems classification when the samples have the classes imbalanced

Chiu *et al.*^[13] applied PCA for feature reduction and MLP to extract characteristics. The authors used *k*-fold cross-validation and obtained 86.97% accuracy. The breast cancer dataset was used with nine attributes. Among them, five important attributes are selected for further prediction. Various other models are also compared and found that MLP outperforms better. The paper did not focus on continuous learning and training the network. Motarwar *et al.*^[23] applied different machine-learning techniques to predict heart disease. The dataset is selected from Cleveland and is preprocessed. Feature selection is applied for selecting features for training. Among the ML algorithms, RF performed better. Atallah and Al-Mousa^[14] focused on predicting

heart disease using different machine-learning methods. Among the various ML algorithms, the author achieved 90% from the ensemble model. But the paper did not focus on any feature reduction or selection process. Mohit *et al.*^[15] took heart disease, diabetes, breast cancer for analysis. They developed a web application to get the disease symptoms and prediction is done based on the symptoms. SVM, LR, and KNN were involved in this study and found heart disease-KNN in 84%, breast cancer LR in 95%, and diabetes LR in 77%. Like other papers, the author has focused on heart, breast cancer, and diabetes. They should focus on diseases that can be solved by deep learning techniques.

Dhomse Kanchan *et al.*^[16] focused on predicting heart disease. PCA has been used for feature selection and used diabetic information of the patient to predict heart disease. Algorithms like DTs, SVM, and Naïve Bayes have been applied and found that SVM performed better. Optimization techniques were not used in this paper. Also, other ML algorithms were not applied for the study. Ahmed *et al.*^[17] worked on finding whether the patient is having diabetes disease. ANN and SVM are integrated by fuzzy logic and used for predicting diabetes disease. The model achieved 94.87% accuracy. Preprocessing of data is carried out but not used any feature selection or extraction techniques. Keniya *et al.*^[56] applied the KNN algorithm for around 230 diseases and tried to predict the based on the symptoms of the patients. The proposed model involved kernel Naïve Bayes, DT, subspace KNN, and RusBoost algorithm. The experiment reached 93.5% accuracy by weighted KNN. But the author did not concentrate on preprocessing and feature selection techniques which will increase the computational time. Kumari and Mehta^[19] tried to use various machine-learning techniques to improve the accuracy. Among them, LDA performed better with an accuracy of more than 80%. The paper did focus on enhancing the performance using a voting ensemble and then the weakest performance classifier is selected. The author did not concentrate on optimizing the dataset. Jatav and Sharma^[20] applied SVM, RF to the dataset of different diseases like diabetes, liver, and kidney disease. The model performed better with these algorithms. The paper did not focus on data mining techniques and optimization of the dataset.

In November 2019 to diagnose COVID-19, the RT-PCR technique was the standard method. But this method has drawbacks like less sensitivity and needs more time to produce the result. Also, this method does not produce false positives but there may be around 30–35% of false negatives. To overcome RT-PCR drawbacks, there came analysis of the X-ray images, computed tomography, and scanned reports which will increase the speed of diagnosing COVID-19. The current and

standard technique used for selecting the specimen from the upper and lower respiratory of a patient is RT-PCR. The RNA Strands will be extracted from the collected specimen. But this method has many limitations. This test has to be carried out in the laboratory by a specialist and it is expensive. The RT-PCR test needs 4–5 h for a single specimen. To overcome the limitations in RT-PCR, the CT scan is widely used but they do not confirm the diagnosis. Also, patients will be exposed to unwanted radiations due to CT scans. The statistical, mathematical, machine learning under AI is also very helpful for the healthcare experts in diagnosing COVID-19 efficiently to bring the disease to control.^[57]

Machine learning is capable of deriving the prediction model without any previous knowledge. The model is efficient to develop huge patterns from data that are complex and noisy. At present, ML techniques like SVM, RF, K-means, and LR are used by many researchers for predicting the virus. The most important and effective algorithm is RF which utilizes many trees to predict the samples. Thus, all the problems in the COVID-19 infection RF play a major role. In the classification and regression of data, the SVM is very useful because of its performance.^[58] The models that have been used in the papers are SVM, LR, PR, ANN, DT, KNN, MLP, EGB, RF, and AdaBoost.

Alotaibi *et al.*^[24] used ANN, SVM, and RF for predicting predict the severity based on the performance of the algorithms, and better performance is achieved by RF. Akhtar *et al.*^[59] tried to predict the disease severity among patients using ML algorithms and 90% of the precision score for predicting the severity. Patel *et al.*^[25] utilized ML algorithms to predict COVID-19 disease severity and found RF performed better with 80% AUC. Gull *et al.*^[26] performed the prediction of severity in COVID patients with the help of ML classification techniques and the data-driven technique has higher accuracy. Tuli *et al.*^[27] used ML-based improved model to predict the COVID-19 possible threat across the world; they found that the data-driven technique has higher accuracy. Ardabili *et al.*^[28] compared between ML and soft computing models to predict the sudden occurrence of COVID-19 and results of MLP and ANFIS has better accuracy. Iwendi *et al.*^[29] performed RF technique with AdaBoost algorithm to predict the COVID-19 virus severity and there is a possibility of recovery or death where 94% accuracy has been achieved by this model.

Aljameel *et al.*^[30] applied LR, RF, and XGB in their model to predict the COVID-19 disease in patients in which among the classification algorithm, RF performed well with 95% accuracy. Naeem *et al.*^[60] applied ARDL and compared SVM, RF, KNN, and ANN for predicting the outbreak of COVID-19 in which the results in predicting

the COVID-19 the AI-based model have high accuracy. Abdullha and Abujar^[61] used LR and KNN in their model to analyze the data and predict the situation in the future and found KNN obtained higher accuracy of 99%. Farooq and Bazaz^[62] used ANN to model and forecast the COVID-19 virus in the most affected parts of India and the model is intellectually able to forecast the spread of the disease. Jain *et al.*^[31] applied ML models and Ensemble learning techniques to predict the SARS-CoV, and CoV-2 using ML models and ensemble techniques with a B-cell dataset, the higher accuracy is obtained by ensemble learning techniques. Traditional ML models and federated learning models were used by the author^[32] to predict the infection caused by COVID-19 and its recovery rate using chest X-ray. SoftMax activation function and SGD optimizer give better accuracy.

Khakharia *et al.*^[33] applied ML models to predict the COVID-19 outbreak in 10 countries in which 99.93% accuracy was obtained by ARMA. ARIMA, SVR, NN, and LR have been used to predict the length of COVID-19 cases across India to decrease the spread of COVID-19; preventive measures have to be followed.^[34] SVM, ANN, RF, DT, LR, and KNN have been used to predict the mortality risk in COVID-19 patients and the overall accuracy of 89.98% was achieved by the model.^[35] The authors^[36] used algorithms like LR, DT, SVM, NB, and ANN for predicting the infection of COVID-19 with the help of an epidemiology dataset where DT has higher accuracy of 94.99%. Elman neural network and SVM algorithms were used to predict the confirmed cases, death, and recovered cases and found SVM along with fuzzy granulation performs well in the prediction.^[37] ML algorithms have been used to predict the COVID-19 virus to identify the symptoms and the result has high accuracy for the dataset taken.^[38] LR, MLP, and vector autoregression were applied to predict the spread of COVID-19 where MLP provides better results compared to other methods.^[39]

SVM algorithms are used to forecast and predict the COVID-19 outbreak,^[40] and compared to the neural network, SVM performed better with 98.88% accuracy. SSLPNN and Gaussian process regression is applied to classify and predict the COVID-19 confirmed cases where the SSLPNN performed better than the Gaussian process regression.^[17] The author^[63] used lower absolute reduction and selection operator, SVM, and LR to predict the individuals affected by COVID-19, but in predicting the new corona death, the linear regression is effective. LR, MLP, RF, and SVM are used by the author^[41] to achieve the disease curve, the tendency of the epidemic. The performance of the algorithms is compared and found that SVM is effective. SVM, rough set, Bayesian ridge, polynomial regression, RNN, and the SIR model were used to predict the recovery rate and pandemic

in which RNN was most effective.^[42] SVM is applied to predict the COVID-19 spread and analyze its growth rate and the method is more effective and accurate.^[43] LR, SVM, PR, and LSTM algorithms were used to predict the rate of growth in COVID-19 and found that the sigmoid model produces higher accuracy.^[44]

LR, SVM, LASSO, and ES techniques were applied to forecast the COVID-19 factors that threaten and a better result was achieved by ES.^[45] RF, linear model, SVM, and NN methods are used to predict the death, recovered and confirmed cases of COVID-19, and RF performed better.^[46] Ensemble ML, BOW, TF/IDF, LR, and MNB were applied to predict COVID where MNB and LR showed better results.^[47] LR, PR, and Holts method have been utilized to predict the cases of COVID-19 in the future and the methods obtained a 0.99 R^2 value.^[48] The polynomial regression method is used to predict the COVID-19 outbreak in which PR predicts the new cases better.^[49]

Ghani *et al.*^[64] researched predicting breast cancer using KNN, ANN, NB, and DT as classification algorithms for the dataset taken from blood samples. After applying recursive elimination of attributes, the author obtained that age, resistin, glucose, BMI, and HOMA are selected for classification. The experimental results produced 80% accuracy for ANN. The feature selection was not compared and analyzed. Also, they concentrated only on breast cancer. For the same breast cancer dataset, SVM, NB, and DT C4.5 were applied by Pritom *et al.*^[65] and used a feature selection algorithm. They found that the algorithm has improved the accuracy after applying ranker. Also, in the research work of Kumar *et al.*^[66] SVM produced greater accuracy in the WEKA software. In another paper, NB, DT, J48, and simple logistics were used for predicting breast cancer, whereas in Rawal,^[67] for the same breast cancer data, the paper was divided into three domains as prediction, diagnosis, and treatment. The model used SVM, KNN, RF, and LR. The time taken to build the model was also compared and found that SVM took less time. SVM achieved better accuracy of 97.1%. The paper did not utilize feature selection and reduction techniques.

Jangle and Narayankar^[51] and Gultepe and Rashed^[68] both researched predicting heart disease using machine learning, namely, DT, NB, J48, and voted DT for classification. PCA was applied as feature selection by both. The authors found that the DT and J48 achieved higher accuracy. For the same heart disease data, Shafique *et al.*^[69] used DT, NB, and a neural network for classification and found Naïve Bayes performed better. The paper used the WEKA tool for feature selection. But this model did not concentrate on other algorithms. The algorithm parameters were not changed in this model.

So, by changing the parameters, the data can be used for designing a system on knowledge.

From Table 1, it is clearly depicted that the list of journals that have dealt with computational intelligence approaches and the results obtained from the models. The comparison of different approaches is showcased in the below graph. From Figure 3, we can infer that few machine-learning algorithms outperform better compared to other approaches. Specifically, RF, Naïve Bayes, LR, ARMA, and SVM reach a better accuracy in predicting the disease.

Optimization Techniques

In this section, the papers that used various optimization techniques for predicting the disease and reducing the features are handled. The algorithms like whale optimization, GA, swarm optimization, and DA are applied for predicting the disease. All these algorithms are used for optimizing the features of the dataset to reduce the complexity. A GA is an evolutionary technique utilized for optimization. PSO is capable of optimizing using a population candidate solution.^[70] It is mainly used in the clustering and watermarking of images.

Tamilselvi and Kumar^[71] tried to predict the disease with the help of improved whale optimization along with a trapezoidal neural network using the UCI dataset of patients. These two methods were used and the efficacy is related to other techniques, namely, the GA and ant colony method. They found that their model produced an accuracy of 86% and exhibited that their model is better. Jabbar *et al.*^[72] applied KNN along with a GA to classify the disease with the UCI dataset of heart patients. They have stated that their model is competitive with another model with high accuracy of 95.73% which is 5% less when it is compared to accuracy without the usage of GA. KNN and GA are applied over the attribute subset which is used to maximize the accuracy. The main drawback of their model is that it is not suitable for breast cancer and tumor.

Islam *et al.*^[73] produced a model that uses PCA for reducing the features along with k means for clustering and hybrid GA. The hybrid GA is used to solve clustering issues. This model has been applied to the heart disease dataset which obtained 94.06% accuracy. They also state that combinatorial optimization has been improved by using k means with GA. Akhil Jabbar *et al.*^[74] developed a model for predicting heart disease with the help of associative classification and a GA. The advantage of using GA is to identify prediction rules that have high accuracy and values. The model achieved an accuracy of 92.8% which is higher when compared with models

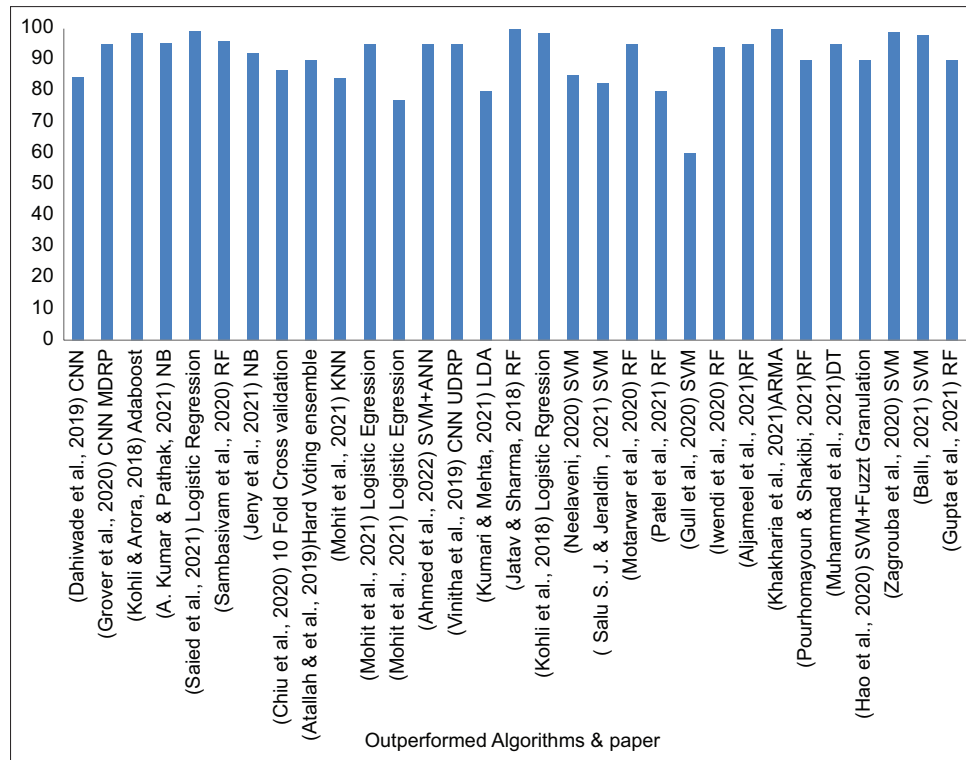


Figure 3: Comparison of computational intelligence algorithms from various papers

that do not use GA. Husin *et al.*^[75] developed a model for the prediction of dengue disease with the help of GA and neural networks and proved that their prediction model is better when compared with other models that are standalone. Kaur and Singh^[76] implemented the GA for predicting heart disease that has 14 features. It has an accuracy of 73.46%. The major objective of this study is to reduce the cost of treatment by predicting the disease and its treatment early. Asghar *et al.*^[77] focused on developing a model, namely, a convolutional neural network. This model helps in training and testing the data. PSO is used for selecting the features from the dataset that are important for constructing the model. After performing PSO, a classification method, namely, SVM is applied. By applying this model, the author found that this model determines 99.81% accuracy which is quick and reliable compared to the regular RT-PCR.

Helal *et al.*^[78] made a study on the medical dataset to classify based on the algorithm and also the author used optimization techniques to reduce the features in the process of predicting cancer patients and diabetes. The authors found that RF performs better when compared with KNN and Naïve Bayes. The major drawback of this system is that pleasing accuracy has not been achieved by the algorithms for predicting diabetes patients. Kanwal *et al.*^[79] applied GA on the UCI medical dataset of heart disease is used for selecting the features. Those selected attributes are passed to the ML algorithms like SVM, NN, NB, DL, and LR. They found that the model has

obtained 92% accuracy in predicting the disease. Among the proposed algorithms, Naïve Bayes and LR performed better. Murthy and Meenakshi^[80] developed a model for heart disease prediction using a neurogenetic approach. A GA is used for selected the feature subset. This subset is useful in angiographic-level prediction. The authors used the Cleveland clinic dataset with 303 patients and the model achieved 89.5% accuracy. Shichkina *et al.*^[81] used neural network and GA for processing the data collected using phone calls of patients affected by Parkinson's disease. The result achieved has better accuracy. Mienye and Sun^[82] proposed a deep-learning technique for predicting the patients' suffering from heart disease. To perform the optimization in parameters, PSO is applied for tuning the stacked sparse auto-encoder. This method efficiently predicted heart disease with 97.3% accuracy on the Cleveland heart dataset.

Pandiaraj *et al.*^[83] demonstrated an experiment to predict the patients with heart disease. The author used a GA and SVM in this process. They also used this method for analyzing the diabetes patients who come with heart disease. Johnson *et al.*^[84] performed diagnosis and assessment in Alzheimer's patients. In this process, LR and GAs are used to identify the risk in patients. They found that to achieve better performance, it is important to optimize the large set of variables to a smaller set. That smaller set contains very important attributes. Suvarna *et al.*^[85] used PSO algorithm to find a solution for predicting heart disease. They used a dataset from UCI

with 14 attributes. They found that the proposed method outperforms the traditional optimizing algorithm. Yaseen *et al.*^[86] used DADA along with ABC techniques for optimizing the weights between the ANN network. The algorithms are compared and found that ANN along with DA performs better than the other algorithms. The algorithms are tested with three different datasets with patients suffering from Cleveland heart, diabetes, and kidney disease. The optimization techniques were discussed in detail and found that these techniques help the model to reduce the number of features in the dataset. These algorithms optimize the dataset and identify the important attributes that are needed for further classification. The accuracy has been increased by using the optimization techniques when compared to applying feature selection and feature reduction techniques.

Conclusion

This paper has empirically analyzed many computational intelligence techniques and tactics that have been used by many researchers in predicting various diseases. This study mainly concentrated on different machine-learning algorithms that are used in predicting different diseases. Depending on the data on clinical disease, the models have been developed. These models can be used to improve the knowledge about predicting the disease with various techniques in machine learning. To understand these models, it is very clear that a large amount of data is required. The data used by these models are mostly clinical text. The most powerful technique that can be used to diagnose the disease precisely could be machine learning. All these techniques that are studied in this review can be used as a dependable tool to act against various diseases. Different machine-learning algorithms have been utilized by various researchers to diagnose the disease and increase the accuracy in predicting the disease. Among them, Naïve Bayes achieved 95.21%, LR 98.97%, and RF with 99.35%, and SVM with 98.8% for COVID dataset accuracy. The algorithms like whale optimization, GA, swarm optimization, and DA are applied for optimizing the dataset and predicting the disease. All these algorithms are used for optimizing the features of the dataset to reduce the complexity. After comparing these optimizing algorithms, genetic algorithms work better.

Acknowledgments

This study was approved by VIT University, Vellore. The authors would like to thank all the professors who helped in the outcome of review paper.

Author biography

Geetha Narasimhan is currently studying Ph.D. and working as an Assistant Professor, School of Computer Science and Engineering at Vellore Institute

of Technology, Vellore. She completed M. Tech in Computer Science and Engineering at VIT, Vellore. She has 7 years of experience in teaching field. She published many papers and book chapters in reputed journals. Her area of interest is machine learning, deep learning, soft computing, optimization, and data mining. She has 7 years of experience in teaching field. She published many papers and book chapters in reputed journals. Her area of interest is machine learning, deep learning, soft computing, optimization, and data mining.

Dr. Akila Victor, Akila Victor is currently working as an Associate Professor in School of Computer Science and Engineering in Vellore Institute of Technology, Vellore. She has completed Ph.D. at VIT University, Vellore, India in 2018. Her research interest includes image processing, machine learning, medical image processing, computer vision, object detection, infrared imaging, and machine learning. She has published more than 15 research papers in reputed international journals and book chapters. She is a life member of CSI.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Orbann C, Sattenspiel L, Miller E, Dimka J. Defining epidemics in computer simulation models: How do definitions influence conclusions? *Epidemics* 2017;19:24-32.
2. Satcher D. U. S. Department of Health and Human Services Public Health Service. 2010. Retrieved March 22, 2023. Available from: <https://www.cdc.gov/nchs/data/hus/hus94.pdf>.
3. Kalantari A, Kamsin A, Shamshirband S, Gani A, Alinejad-Rokny H, Chronopoulos AT. Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions. *Neurocomputing* 2018;276:2-22.
4. Dahiwade D, Patle G, Meshram E. Designing disease prediction model using machine learning approach. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) IEEE; 2019. p. 1211-5.
5. Grover A, Kalani A, Dubey SK. Analytical approach towards prediction of diseases using machine learning algorithms. In 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence) IEEE; 2020. p. 793-7.
6. Kohli PS, Arora S. Application of machine learning in disease prediction. In 2018 4th International conference on computing communication and automation (ICCCA) IEEE; 2018. p. 1-4.
7. Kumar A, Pathak MA. A machine learning model for early prediction of multiple diseases to cure lives. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 2021;12:4013-23.
8. Saied IM, Arslan T, Chandran S. Classification of Alzheimer's disease using RF signals and machine learning. *IEEE J Electromagn RF Microw Med Biol* 2021;6:77-85.
9. Kumar N, Nripendra Narayan Das DG, Gupta K, Bindra J. Learning models. *J Healthc Eng* 2010;6:63-83.
10. Syed AH, Khan T, Hassan A, Alromema NA, Binsawad M, Alsayed AO. An ensemble-learning based application to predict

- the earlier stages of Alzheimer's disease (AD). *IEEE Access* 2020;8:222126-43.
11. Sambasivam G, Amudhavel J, Sathya G. A predictive performance analysis of vitamin D deficiency severity using machine learning methods. *IEEE Access* 2020;8:109492-507.
 12. Jeny J, Reddy N, Aishwarya P, Samreen. A Classification Approach for Heart Disease Diagnosis using Machine Learning. 2021. p. 456-9. 10.1109/ISPC53510.2021.9609468.
 13. Chiu HJ, Li THS, Kuo PH. Breast cancer-detection system using PCA, multilayer perceptron, transfer learning, and support vector machine. *IEEE Access* 2020;8:204309-24.
 14. Atallah R, Al-Mousa A. Heart disease detection using machine learning majority voting ensemble method. In 2019 2nd international conference on new trends in computing sciences (ictcs) IEEE; 2019. p. 1-6.
 15. Mohit I, Kumar KS, Reddy AUK, Kumar BS. An approach to detect multiple diseases using machine learning algorithm. *J Phys Conf Ser* 2021;2089. doi: 10.1088/1742-6596/2089/1/012009.
 16. Kanchan BD, Kishor MM. Study of machine learning algorithms for special disease prediction using principal of component analysis. In 2016 international conference on global trends in signal processing, information computing and communication (ICGTSPICC). IEEE; 2016. p. 5-10.
 17. Ahmed U, Issa GF, Khan MA, Aftab S, Khan MF, Said RAT, *et al.* Prediction of diabetes empowered with fused machine learning. *IEEE Access* 2022;10:8529-38.
 18. Vinitha S, Sweetlin S, Vinusha H, Sajini S. Disease prediction using machine learning over big data. *Comput Sci Eng Int J* 2019;7:556-9.
 19. Kumari A, Mehta AK. A novel approach for prediction of heart disease using machine learning algorithms. In 2021 Asian Conference on Innovation in Technology (ASIANCON), IEEE; 2021. p. 1-5.
 20. Jatav S, Sharma V. An algorithm for predictive data mining approach in medical diagnosis. *Int J Comput Sci Information Technol* 2018;10:11-20.
 21. Neelaveni J, Devasana MG. Alzheimer disease prediction using machine learning algorithms. In 2020 6th international conference on advanced computing and communication systems (ICACCS) IEEE; 2020. p. 101-4.
 22. Alphonsa Salu SJ, Jeraldin Auxillia D. Detection of knee joint disorders using SVM classifier. *Int J Sci Res Sci Technol* 2021;8:261-71.
 23. Motarwar P, Duraphe A, Suganya G, Mariappan P. Cognitive Approach for Heart Disease Prediction using Machine Learning. 2020. p. 1-5. 10.1109/ic-ETITE47903.2020.242.
 24. Alotaibi A, Mohammad S, Alshahrani A. Prediction of severity of COVID-19-infected patients using. *Computers* 2021;10:31.
 25. Patel D, Kher V, Desai B, Lei X, Cen S, Nanda N, *et al.* Machine learning based predictors for COVID-19 disease severity. *Sci Rep* 2021;11:1-7. doi: 10.1038/s41598-021-83967-7.
 26. Gull H, Krishna G, Aldossary M, Iqbal S. Severity Prediction of COVID-19 Patients Using Machine Learning Classification Algorithms: A Case Study of Small City in Pakistan with Minimal Health Facility. 2020. p. 1537-41. 10.1109/ICCC51575.2020.9344984.
 27. Tuli S, Tuli S, Tuli R, Gill SS. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. *Internet Things* 2020;11:100222. doi: 10.1016/j.iot. 2020.100222.
 28. ArdabiliSF, Mosavi A, GhamisiP, FerdinandF, Varkonyi-KoczyAR, Reuter U, *et al.* COVID-19 outbreak prediction with machine learning. *MedRxiv* 2020. doi: 10.1101/2020.04.17.20070094.
 29. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, *et al.* COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health* 2020;8:1-9. doi: 10.3389/fpubh. 2020.00357.
 30. Aljameel SS, Khan IU, Aslam N, Aljabri M, Alsulmi ES. Machine learning-based model to predict the disease severity and outcome in COVID-19 patients. *Sci Program* 2021. doi: 10.1155/2021/5587188.
 31. Jain N, Jhunjhara S, Garg H, Gupta V, Mohan S, Ahmadian A, *et al.* Prediction modelling of COVID using machine learning methods from B-cell dataset. *Results Physics* 2021;21:103813. doi: 10.1016/j.rinp.2021.103813.
 32. Abdul Salam M, Taha S, Ramadan M. COVID-19 detection using federated machine learning. *PLoS One* 2021;16:e0252573. doi: 10.1371/journal.pone.0252573.
 33. Khakharia A, Shah V, Jain S, Shah J, Tiwari A, Daphal P, Warang M, Mehendale N. Outbreak prediction of COVID-19 for dense and populated countries using machine learning. *Ann Data Sci* 2021;8:1-19. doi: 10.1007/s40745-020-00314-9.
 34. Ogundokun RO, Awotunde JB. Machine learning prediction for covid 19 pandemic in India. *medRxiv* 2020. p. 2020-05.
 35. Pourhomayoun M, Shakibi M. Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Health* 2021;20:100178. doi: 10.1016/j.smhl.2020.100178.
 36. Muhammad LJ, Algehyne EA, Usman SS, Ahmad A, Chakraborty C, Mohammed IA. Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset. *SN Comput Sci* 2021;2:1-13. doi: 10.1007/s42979-020-00394-7.
 37. Hao Y, Xu T, Hu H, Wang P, Id YB. Prediction and analysis of Corona Virus Disease 2019. *PLoS One* 2020;15:e0239960. doi: 10.1371/journal.pone.0239960.
 38. Zoabi Y, Deri-Rozov S, Shomron N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *Npj Digital Med* 2021;4:1-5. doi: 10.1038/s41746-020-00372-6.
 39. Sujath R, Chatterjee JM, Hassanien AE. A machine learning forecasting model for COVID-19 pandemic in India. *Stoch Environ Res Risk Assess* 2020;34:959-72.
 40. Zagrouba R, Khan MA, Atta-Ur-Rahman Saleem MA, Mushtaq MF, Rehman A, Khan MF. Modelling and simulation of COVID-19 outbreak prediction using supervised machine learning. *COMPUTERS MATERIALS & CONTINUA.* 2020;66:2397-407.
 41. Balli S. Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods. *Chaos Solitons Fractals* 2021;142:110512. doi: 10.1016/j.chaos. 2020.110512.
 42. Arun SS, Iyer GN. On the analysis of COVID19-novel corona viral disease pandemic spread data using machine learning techniques. In 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS) IEEE; 2020. p. 1222-7.
 43. Yadav M, Perumal M, Srinivas M. Analysis on novel coronavirus (COVID-19) using machine learning methods. *Chaos Solitons Fractals* 2020;139:110050. doi: 10.1016/j.chaos. 2020.110050.
 44. Sengupta S, Mugde S, Sharma G. Covid-19 pandemic data analysis and forecasting using machine learning algorithms. *MedRxiv* 2020. doi: 10.1101/2020.06.25.20140004.
 45. Rustam F, Reshi AA, Mehmood A, Ullah S, On BW, Aslam W, *et al.* COVID-19 future forecasting using supervised machine learning models. *IEEE Access* 2020;8:101489-99.
 46. Gupta VK, Gupta A, Kumar D, Sardana A. Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model. *Big Data Mining and Analytics*, 2021;4:116-23.
 47. Khanday AMUD, Rabani ST, Khan QR, Rouf N, Mohi Ud Din M. Machine learning based approaches for detecting COVID-19 using clinical text data. *Int J Inf Technol* 2020;12:731-9.
 48. Lakshmanarao A, Seshadri Rao C, Gadde S. Analyzing and predicting Covid-19 epidemic using machine learning techniques. *IOP Conf Ser Mater Sci Eng* 2021;1074:012018. doi: 10.1088/1757-899X/1074/1/012018.
 49. Li L, Yang Z, Dang Z, Meng C, Huang J, Meng H, *et al.* Propagation analysis and prediction of the COVID-19. *Infect Dis Model* 2020;5:282-92.
 50. Elmrabbit N, Zhou F, Li F, Zhou H. Evaluation of Machine Learning Algorithms for Anomaly Detection. 2020. p. 1-8. 10.1109/

- CyberSecurity49315.2020.9138871.
51. Jangle P, Narayankar S. Alternating decision trees for early diagnosis of heart disease. *Int J Eng Comput Sci* 2016;5:16070-2.
 52. Chaurasia V, Pal S, Tiwari BB. Chronic kidney disease: A predictive model using decision tree. *Int J Eng Res Technol* 2018;11:1781-94.
 53. Vikas C, Saurabh P. Data mining approach to detect heart diseases. *Int J Advanced Computer Science and Information Technology*. 2013;2:56-66.
 54. Chaurasia V, Pal S, Tiwari BB. Prediction of benign and malignant breast cancer using data mining techniques. *J Algorithm Computat Technol* 2018;12:119-26.
 55. Williams R, Shongwe T, Hasan AN, Rameshar V. Heart disease prediction using machine learning techniques. In 2021 international conference on data analytics for business and industry (ICDABI). IEEE; 2021. p. 118-23.
 56. Keniya R, Khakharia A, Shah V, Gada V, Manjalkar R, Thaker T, *et al*. Disease prediction from various symptoms using machine learning. *Social Science Research Network Electron J* 2020. doi: 10.2139/ssrn.3661426.
 57. Xu X, Jiang X, Ma C, Du P, Li X, Lv S, *et al*. A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering* 2020;6:1122-9.
 58. Swapnarekha H, Behera HS, Nayak J, Naik B. Role of intelligent computing in COVID-19 prognosis: A state-of-the-art review. *Chaos Solitons Fractals* 2020;138:109947. doi: 10.1016/j.chaos.2020.109947.
 59. Akhtar M, Kraemer MUG, Gardner LM. A dynamic neural network model for predicting risk of Zika in real time. *BMC Med* 2019;17:171.
 60. Naeem M, Yu J, Aamir M, Khan SA, Adeleye O, Khan Z. Comparative analysis of machine learning approaches to analyze and predict the COVID-19 outbreak. *PeerJ Computer Science* 2021;7:e746.
 61. Abdullha A, Abujar S. COVID-19: data analysis and the situation prediction using machine learning based on Bangladesh perspective. In 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (ISAI-NLP). IEEE; 2020. p. 1-8.
 62. Farooq J, Bazaz MA. A deep learning algorithm for modeling and forecasting of COVID-19 in five worst affected states of India. *Alexandria Eng J* 2021;60:587-96.
 63. Mojada RK, Yadav A, Prabhu AV, Natarajan Y. Machine learning models for covid-19 future forecasting. *Mater Today Proc* 2020. doi: 10.1016/j.matpr.2020.10.962.
 64. Abdullha A, Abujar S. COVID-19: data analysis and the situation prediction using machine learning based on Bangladesh perspective. In 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (ISAI-NLP). IEEE; 2020. p. 1-8.
 65. Pritom AI, Munshi MAR, Sabab SA, Shihab S. Predicting breast cancer recurrence using effective classification and feature selection technique. 19th International Conference on Computer and Information Technology, ICCIT, Dhaka, Bangladesh, 2016. p. 310-14. doi: 10.1109/ICCITECHN.2016.7860215.
 66. Kumar GR, Ramachandra GA, Nagamani K. An efficient prediction of breast cancer data using data mining techniques. *Int J Innov Eng Technol* 2013;2:139-44.
 67. Rawal R. Breast cancer prediction using machine learning. *Appl Comput Syst* 2020;25:163-71.
 68. Gultepe Y, Rashed S. The use of data mining techniques in heart disease prediction. *Int J Comput Sci Mobile Comput* 2019;8:136-41.
 69. Shafique U, Majeed F, Qaiser H, Mustafa IU. Data mining in healthcare for heart diseases. *Int J Innov Appl Stud* 2016;10:1312-22.
 70. Kennedy J, Eberhart R. Particle swarm optimization. In Proceedings of ICNN'95-international conference on neural networks. IEEE; 1995;4:1942-8.
 71. Tamilselvi K, Kumar DKR. An efficient disease prediction in big data using neural network based optimization method. *Turkish J Comput Math Educ* 2021;12:1140-7.
 72. Jabbar MA, Deekshatulu BL, Chandra P. Heart disease prediction using lazy associative classification. In 2013 International Multi-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s). IEEE; 2013. p. 40-6.
 73. Islam MT, Rafa SR, Kibria MG. Early prediction of heart disease using PCA and hybrid genetic algorithm with k-means. In 2020 23rd International Conference on Computer and Information Technology (ICCIT). IEEE; 2020. p. 1-6.
 74. Jabbar, M. A., Deekshatulu, B. L., and Chandra, P. Heart Disease Prediction System using Associative Classification and Genetic Algorithm. 2013. *ArXiv./abs/1303*.
 75. Husin NA, Mustapha N, Sulaiman MN, Yaakob R. A hybrid model using genetic algorithm and neural network for predicting dengue outbreak. In: 2012 4th Conference on Data Mining and Optimization (DMO). Langkawi, Malaysia: IEEE; 2012. p. 23-7. doi: 10.1109/DMO.2012.6329793.
 76. Kaur B, Singh W. Analysis of heart attack prediction system using genetic algorithm. *Int J Adv Technol Eng Sci* 2015;87-94.
 77. Asghar MA, Razzaq S, Rasheed S. A robust technique for detecting sars-cov-2 from x-ray image using 2d convolutional neural network and particle swarm optimization. In 2020 14th International Conference on Open Source Systems and Technologies (ICOSST). IEEE; 2020. p. 1-6.
 78. Al Helal M, Chowdhury AI, Islam A, Ahmed E, Mahmud MS, Hossain S. An optimization approach to improve classification performance in cancer and diabetes prediction. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE). IEEE; 2019. p. 1-5.
 79. Kanwal S, Rashid J, Nisar M W, Kim J, Hussain A. An effective classification algorithm for heart disease prediction with genetic algorithm for feature selection. In 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC). IEEE; 2021. p. 1-6.
 80. Murthy H N, Meenakshi M. Dimensionality reduction using neuro-genetic approach for early prediction of coronary heart disease. In International conference on circuits, communication, control and computing. IEEE; 2014. p. 329-32.
 81. Shichkina Y, Irishina Y, Stanevich E, de Jesus Plasencia Salgueiro A. Application of genetic algorithms for the selection of neural network architecture in the monitoring system for patients with Parkinson's disease. *Appl Sci* 2021;11:5470. doi: 10.3390/app1125470.
 82. Mienye ID, Sun Y. Improved heart disease prediction using particle swarm optimization based stacked sparse autoencoder. *Electron* 2021;10:2347. doi: 10.3390/electronics10192347.
 83. Annamalai, Pandiaraj & Sanmugaraja, Lakshmanaparakash & Kanna, P Rajesh. Effective Heart Disease Prediction Using Hybridmachine Learning. 731-738. 10.1109/ICICV50876.2021.9388635. 2021.
 84. Johnson P, Vandewater L, Wilson W, Maruff P, Savage G, Graham P, *et al*. Genetic algorithm with logistic regression for prediction of progression to Alzheimer's disease. *BMC Bioinformatics* 2014;15(Suppl 16):S11. doi: 10.1186/1471-2105-15-S16-S11.
 85. Suvarna C, Sali A, Salmani S. Efficient heart disease prediction system using optimization technique. In 2017 International Conference on Computing Methodologies and Communication (ICCMC). IEEE; 2017. p. 374-9.
 86. Yasen M, Al-Madi N, Obeid N. Optimizing neural networks using dragonfly algorithm for medical prediction. In 2018 8th international conference on computer science and information technology (CSIT). IEEE; 2018. p. 71-6.