

A systematic review of the machine learning algorithms for the computational analysis in different domains

Ravita Chahar^{1*} and Deepinder Kaur²

Department of Computer Science, Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India¹

Department of Computer Science, Chandigarh University, Punjab, India²

Received: 31-July-2020; Revised: 20-October-2020; Accepted: 25-October-2020

©2020 Ravita Chahar and Deepinder Kaur. This is an open access article distributed under the Creative Commons Attribution (CC BY) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In this paper machine learning algorithms have been discussed and analyzed. It has been discussed considering computational aspects in different domains. These algorithms have the capability of building mathematical and analytical model. These models may be helpful in the decision-making process. This paper elaborates the computational analysis in three different ways. The background and analytical aspect have been presented with the learning application in the first phase. In the second phase detail literature has been explored along with the pros and cons of the applied techniques in different domains. Based on the literatures, gap identification and the limitations have been discussed and highlighted in the third phase. Finally, computational analysis has been presented along with the machine learning results in terms of accuracy. The results mainly focus on the exploratory data analysis, domain applicability and the predictive problems. Our systematic analysis shows that the applicability of machine learning is wide and the results may be improved based on these algorithms. It is also inferred from the literature analysis that at the applicability of machine learning algorithm has the capability in the performance improvement. The main methods discussed here are classification and regression trees (CART), logistic regression, naïve Bayes (NB), k-nearest neighbors (KNN), support vector machine (SVM) and decision tree (DT). The domain covered mainly are disease detection, business intelligence, industry automation and sentiment analysis.

Keywords

Machine learning, CART, NB, KNN, SVM, DT, Computational analysis.

1. Introduction

In the current technology era machine learning algorithms have gained immense popularity for their all-round capability and wide applicability in different domains [1–3]. Different domains needed process formation and automation need the collaboration of machine learning techniques along with the computation capabilities for the complex situations [4,5]. The selection of these algorithms in different domains may depend on different factors. It includes data analysis, patterns exploration, attributes impact, influencing parameters, constants and the need of automation [6–9]. The applicability of machine learning algorithms is very wide, including the healthcare system, disease detection and prediction, weather forecasting, time series analysis, fraud detection and business intelligence [10–18].

The major methods applied to most of the problems are classification and regression trees (CART), logistic regression, naïve Bayes (NB), k-nearest neighbors (KNN), support vector machine (SVM) and decision tree (DT) [15–18]. The major domains where machine learning algorithms frequently used are disease detection, business intelligence, industry automation and sentiment analysis [10–18]. The general process of machine learning algorithm is shown in *Figure 1*. The domain of applicability is shown in *Figure 2*.

So, for better understating the machine learning capabilities and methodological analysis based on the computational behavior required. It covers different machine learning algorithms along with the computational focus for the study and analysis.

*Author for correspondence

The objectives of this paper are as follows:

1. To study and analyze the machine algorithms for different domains.
2. To evaluate different performance constraints which may be useful in performance improvement.
3. To analyse the adaptability for exploratory data analysis, domain applicability and the predictive problems.
4. To explore the latest literature for the gap identification.

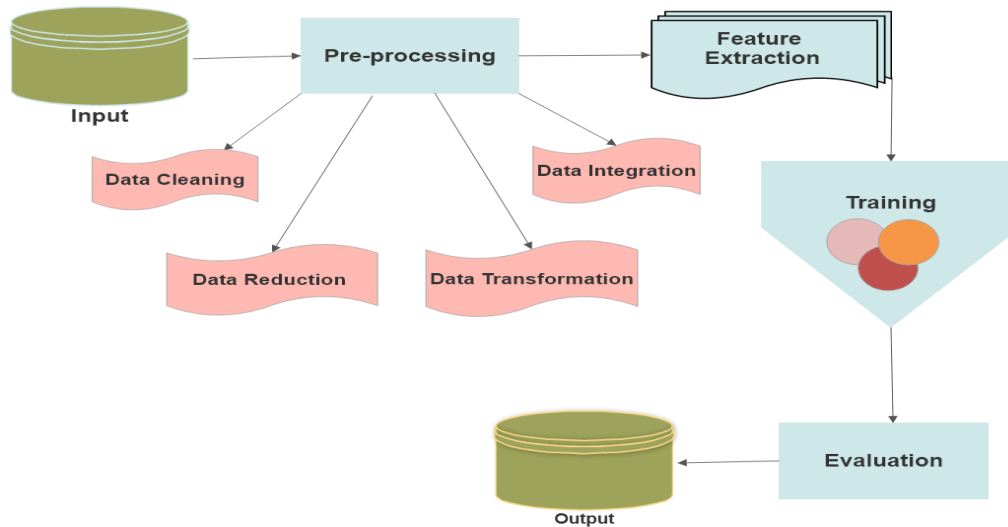


Figure 1 General process of machine learning algorithm

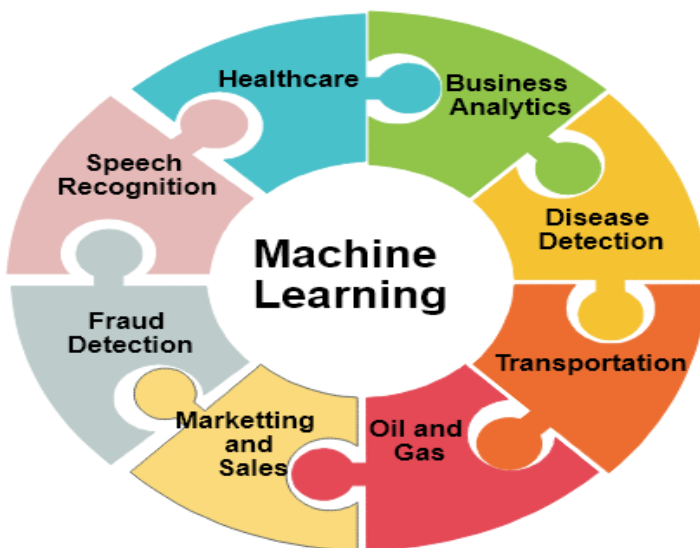


Figure 2 Machine learning uses in different domains

2.Related work

In this section a systematic review and analysis have been presented based on the latest literature focusing on the pros and cons. In 2020, Postnikov et al. [19], proposed a tool for diagnosis the availability of drug in body which is based on special biomarkers taken from blood and are reliable in clinics. They trained

the CatBoost with this clinical data taken from pulmonary tuberculosis. They suggested the use of CatBoost in the medical field also instead of pure engineering. In 2020, Panigrahi et al. [20] discussed about the approach used for detecting the disease in Corn leaves. The approach proposed by the authors is a convolutional neural network (CNN) where they

used the Adam optimizer with activation functions used in rectified linear. The highest accuracy achieved is 98.78% for the detection. This approach is useful for the poor farmers who can save them from the crop loss and gives encouragement to the agricultural system. They proposed to use the given approach with deep learning and machine learning for the identification of more diseases present in corn. In 2020, Jamil et al. [21] proposed a deep learning neural network named CanNet. This neural network is helpful in detecting the breast cancer at an early stage and is proficient to give 99% accuracy in case of sensitivity. They verified various machine learning models with different classifiers and checked the accuracy. A dataset of nine attributes was provided as the input to connect and as an output of this they detected the patient with or without Malignant. In future studies, they suggested to use it on X-ray images for detecting the breast cancer. In 2020, Aruchamy et al. [22], suggested a method to detect the Alzheimer's disease using MRI. They applied different classifier on Grey and White matter because they get affected by Alzheimer's disease. The features extracted from it is the first-order statistical with the help of principal component analysis. If compared with other techniques the result provided by these classifiers are the accuracy of 90.9% in both the white and Grey matter cases with all four classifiers. In 2020, Iadanza et al. [23], applied machine learning to Pupillometric data for detecting the genetic disease in the age of pediatric. They have used Decision support system with the use of Chromatic Pupillometry to assess the functions of inner and outer retinas. They used the SVM classifiers for it. The combinations of these SVMs show the good accuracy performance in sensitivity with the 93.7%. As a futuristic approach they suggested to test the same approach, but with the help of different devices. In 2020, Mane et al. [24], proposed a machine learning model which can detect fourteen different diseases from a single report of chest radiographs. They took the dataset from NIH Clinical center and categorized the disease in 14 classes. They applied deep CNN. They did the comparison of this approach to traditional and found the best result in this proposed method. In 2020, Roopa and Mamatha [25], tried to identify the correlation between Optical Image recognition and machine learning. They found that recurrent neural network and SVM classifiers are more beneficial in detecting and correcting the errors in Optical character recognition text. As a future work they suggested to use CNN of deep learning. In 2020, Majdi et al. [26], proposed a method of classification

done by deep learning in appeared in pathologies in the images of X-ray. They compared their performance with some existing methods and found improved results in an area under a curve in the case of the proposed approach. They found 91% for AUC in case of cardiomegaly and 73% in nodule classification parts. In 2020, Ramírez et al. [27], proposed an approach that can show that auto encoders are helpful in detecting the Parkinson's disease at early stage. They used the Magnetic Resonance Imaging as a data for this identification. They used the both types of spatial autoencoders and Dense variational autoencoders. In 2020, Watanabe et al. [28], used the random forest (RF) as a classifier model for detecting the cervical myelopathy automatically. They computed the hand movement of leap motion and tried to evaluate the automatic performance of detection. Other diseases can be analysed which shows the difficulty in showing the symptoms. This type of technique can be applied to tremor and numbness. In 2020, Ramos-Giraldo et al. [29], identify the stress caused due to drought with the help of machine learning. They applied deep learning and machine learning to the images of crops and then by performing the classification of images they can easily identify the level of crops. With the help of this identification they can easily prepare the schedule. In 2020, Ratley et al. [30], tried to analyze the method to detect the loss in blood in reference to detect the blood cancer with the help of machine learning techniques. They applied it to the both the types of leukemia diseases i.e. with lymphocytic and myeloid. They have discussed various classification techniques after pre-processing techniques, but suggested to use Deep learning techniques because of the large datasets. In 2020, Kundu et al. [31], discussed about the wearable gadgets to detect the health-related issues in the case of disorder, sleep, breath problems etc. They proposed an IoT base predictor for disease identification and then generating an alert system for it. They proposed an architecture for monitoring the health perform the future prediction. They are trying to make it efficient in terms of energy and economics. In 2020, Bini et al. [32] proposed a new approach of using agricultural robots on the fields for sowing the crops. They tried to reduce the labour power. They used the environment images with the help of image processing and then performed pre-processing and extraction for the feature selection with the accuracy of 66.8%. The Kernel function of SVM was used to train the images. In 2020, Kundu et al. [33] tried to focus on the burden faced by physicians faced in endoscopy. Here the authors developed the computer-

aided method for detecting the gastrointestinal diseases. They trained the pixel-labelled images because these types of images are more helpful in detecting the diseases. They performed the Linear discriminate analysis for extracting the images. Finally, they achieve the highest number of accuracies with the pixel images. In 2020, Nidigattu et al. [34] developed a photoplethysmography method for estimating the blood pressure and heart rate with the help of different machine learning model classifiers. RF algorithm found to best in case of accuracy. As a future scope they proposed to use it for hypertension, diabetes etc. In 2020, Tuncer et al. [35], discussed to detect the disease with the help of voice. They proposed as multiclass-pathologic voice with some repetitive features. They used a ternary pattern which is more compact and having multileveled features. The features are finally integrated to MCMTTP for detecting the accuracy in voice. This is the cognitive method and helpful in detecting the level of voice in case of chest infection/disease. In 2020 Iqbal et al. [36], discussed the accurate information required for the detection of disease present in heart with the use of clinical data available. They took the data from the University of California with 14 attributes. They achieved the accuracy of 98.41% with OLM method. They were comparable with other classifiers as well. They suggested to improve this efficiency with a smaller number of attributes.

3.Problem identification

The major problems identified based on the recent survey and analyses are as follows:

- 1) There is the need of specific level-based orientation for proper data extraction and classification.
- 2) There is the need of computational methods along with machine learning aspect of the better performance and complex problems.
- 3) There is the need of data processing which should synchronize the data processing and computation time.
- 4) There is the need of self-adaptive system to automatically modify the run-time behavior of the system based on the priority and requirements.
- 5) As the applicability of machine learning algorithms is domain specific so it should be considered in terms of individual cases and solved separately.

4.Discussion and analysis

Table 1 shows the analysis based on the method and its applicability. Based on the analysis, hybrid model has been found to be better in comparison to all. Based on the Table 2, it has been found that the Gradient Boosting algorithm may be useful for fall detection. DT is found to be better in case of mosquito identification. NB classifier can be used for sentiment classification and also outperforms. SVM has been found to be prominent in several disease detection including pancreatic cancer. CNN gives the best result in detecting the disease in plant leaf. RF after outlier removal gives good result in patients with liver disease. KNN is found to be helpful in the detection of medicinal Plants. RF classifier gives the best accuracy result for predicting the depression, stress and anxiety.

Table 1 Analysis based on the method and its applicability

Sr. no.	Authors	Method	Approach	Domain applied	Single/hybrid
1	Kumari and Jose [37]	The method used in this study was the supervised learning method i.e. SVM.	The SVM was applied for classification of the EEG signal having two different groups i.e. healthy and the dataset of epileptic.	This SVM method is applied in the detection of Seizure in EEG signal with the Analysis of Time Frequency which is also used for diagnosing brain disorders.	Single classifier implementation.
2	Shahbakhti et al. [38]	The pre-processing step was done with the help of principal components analysis. Further, the SVM is used for the	The dataset used by the researcher was of Max Little of the University of Oxford. Initially, the pre-processing step was done by	This study is applied in the detection of Parkinson's disease in the form of speech signal.	This study used combination of PCA and SVM.

Sr. no.	Authors	Method	Approach	Domain applied	Single/hybrid
		classification.	Principal Component Analysis and further the SVM was used for classification.		
3	Bourouhou et al. [39]	This study had used three different classifiers for comparing the efficiency. These classifiers were KNN, NB and SVM.	The data from the UCI was taken. They used the voice recording of patients and healthy people. They did extraction in the features on the voice dataset. Further, three different classifiers were applied on the extracted features.	The researchers had applied all these three classifiers in the identification of Parkinson's diseases.	This paper had the role of SVM, KNN and NB classifiers.
4	Padol and Sawant [40]	The authors adopted the SVM and artificial neural network (ANN) as the classifiers. Further, the fusion of both were used as the new classification techniques.	The researchers used the images and pre-processed those images with the help of K-means clustering. Further, the feature extraction was classified with the help of fusion classification for the identification of diseases.	The different classifiers were applied on the identification of disease in grape leaf.	This study was based on the SVM and ANN.
5	Shetty and Rao [41]	This paper has used the SVM as the classifier on the basis of Machine Learning to detect some specific gait characteristics associated with the Parkinson's disease.	This study took some statistical features from the Gait data. Further, the seven best features were extracted and they were further classified on the basis of function Kernel which was based on SVM.	This study tried to identify the Parkinson's disease with the help of gait analysis.	They used only SVM.
6	Padol and Yadav [42]	The method applied in this paper is SVM with the help of Image Processing. They used the K-mean clustering for the segmentation purpose so that color and texture can be extracted.	The preprocessing was done on the leaf images so that after segmentation feature extraction can be applied. After getting Texture and color of the leaf classification was done with the help of SVM.	This study was based on detecting the disease in grape leaf.	SVM classifier was used.

Sr. no.	Authors	Method	Approach	Domain applied	Single/hybrid
7	Patil et al. [43]	In this study the researchers used the SVM, ANN and RF. They did the comparison among all on the common data set to identify the higher accuracy in results.	The authors pre-processed the leaf image of potato. After feature extraction they performed the classification with the help of different classifiers for the identification of disease. Out of three they found the best result in ANN classifier.	The performance of these three classifiers was applied on disease detection in Potato through images of Leaf.	This study used the three different classifiers named as SVM, ANN and RF.
8	Prakash and Saraswathy [44]	In this paper the researcher used the Image processing technique on images and further SVM was used as the classifier.	The SVM was used as the classifier on the values retrieved from Co-occurrence matrix.	The researchers had applied image processing and support vector machine on the detection of diseases in leaf	SVM was used as the classifier.
9	Islam et al. [45]	The researchers applied the integration of Machine learning and image processing for diagnosing the diseases from the images of leaf. The SVM method was applied with the use of Image segmentation.	This paper has applied SVM classifier as the classification model to extract the features so that it can easily identify the characteristics. This approach helps the farmers to easily get the identification of potato disease.	The SVM classifier model is used with the Image segmentation of Potato leaf. The potato leaves are used for disease identification.	The classifier used in this study was SVM.
10	Masazhar and Kamal [46]	In this paper the method used the Image processing technique along with the multiclass SVM classifier.	The Image processing technique was applied on the images of leaf. After extracting the features, the multiclass SVM classifier was implemented for different classification. Finally, result of 97% accuracy was achieved.	This study is applied on the disease detection for Palm oil leaf.	This paper applied only the SVM as the classifier.
11	Mathew and Anto [47]	In this paper the researchers used the MRI images as the input. Finally, they extract the levels in brain with the help of classification done	This study used the MRI images and applied a diffusion filter which extracted the features and after combining the result of	The SVM classifier method was applied in detecting the Tumour and also perform the classification in images related to	The classifier used was the SVM

Sr. no.	Authors	Method	Approach	Domain applied	Single/hybrid
		by SVM method.	extracted features the SVM classifier was applied. Finally, a SWT decomposed image was received which was used as classifier input.	Brain.	
12	Karan et al. [48]	This paper used the variational mode decomposition for extracting the signals of speech. Further, the SVM classifier was used for classifying the healthy and patients who are suffering from Parkinson's disease.	In this paper speech signals were used to recognize the variations in speech. The speech was further classified in different classes and accordingly they identified the results.	Variational mode decomposition and SVM were used to identify the Parkinson's disease with the help of speech signal.	This paper used the only SVM classifier.
13	Kiss et al. [49]	This paper used the SVM classification with linear kernel. The RoT input was used as the classification method.	In detecting the depression and Parkinson's disease among people speech as a parameter was used. The investigation was done on the Transient parts. Further, The SVM did the classification on three categories like Healthy, depressed and Parkinson's patients.	This SVM classifier was applicable to speech where different categories of patients were identified.	This study used the SVM classifier.
14	Bhimte and Thool [50]	The Image processing and SVM classifier method was used as the disease detection.	The Image processing was done on the cotton leaf to extract the images. These images were further pre-processed and segmented with K-means. Meaningful extracted features were further classified with the help of SVM.	Image processing, k-means clustering and SVM was applied on disease detection in cotton leaf.	Only SVM classifier was used in this study.
15	Kusumo et al. [51]	Image processing techniques were	This study simply applied image	All these different types of machine	This study applied the machine

Sr. no.	Authors	Method	Approach	Domain applied	Single/hybrid
		used to identifying the various features. Various machine learning algorithms like RF, NB, SVM and decision tree were applied	processing on the corn-plant and extracted the features like colour (RGB). They have applied various machine learning algorithms. Among all the algos. The best result achieved was by the SVM.	learning algos. Are applied on the Corn plant.	learning algos. Like RF, NB, DT and SVM.1
16	Win et al. [52]	In this study 7 different machine learning algos. Were used to perform the diagnosis of malignant Mesothelioma disease. They had applied different machine learning algorithms like RF, linear discriminant analysis (LDA), logistic regression (LR), NB, SVM and KNN.	This study performed the experiment on dataset of 324 cases which had 34 features. Seven different types of classifiers were used to compare the accuracy of result and finally the 100% result was achieved by LogR, SVM, RF and DT only.	The comparison of different classifiers was applied on the diagnosis of rare cancer i.e. Malignant Mesothelioma.	This study was applied on seven different types of classifiers i.e. RF, NB, SVM, DT, LogR, KNN and LDA
17	Pardede et al. [53]	In this paper the researcher tried to apply an unsupervised feature machine learning algorithm which uses the convolutional auto encoder for the detection and support-Vector machine as the classifier.	In this paper author applied the convolutional auto encoder which used the handcrafted features. The output of the auto encoder was applied to SVM as the input and further it identified the disease detected in plant.	The convolutional auto encoder and SVM were applied in detecting the disease in plant.	Single: The SVM Classifier is applied in automatic detecting the disease in plant.
18	Gurralla et al. [54]	This paper used the diagnosis method for plant disease. They had applied the SVM Classifier and K-means clustering for image segmentation	The approach followed in this study was the K-means and SVM for image segmentation and image classification.	This study had applied SVM as the classifier on leaf of plant to diagnose the disease on early stage.	Single Model: SVM.
19	Al et al. [55]	This paper tried to perform the text classification on Arabic Lang. Ten different	In this study the researchers took the dataset of different categories like	Various types of supervised classification text were applied on Arabian Text	Hybrid Model: The Ten different classifiers were used in the present study.

Sr. no.	Authors	Method	Approach	Domain applied	Single/hybrid
		classifiers were used by them namely SVM, DT, XGBOOST classifiers, RF, Multi-nominal classifier etc. of the supervised learning.	sports, Technology, Business and applied different types of classifiers on them. Finally, they found the good result in the applicability of SVM in the extracted text features.	Language.	
20	Laksmi and Naik [56]	The author tried to perform feature extraction on the input image and further they applied the SVM and K-nearest neighbour as the classifier for detecting the disease.	This paper considered the different sub-kinds of ALL and AML. They grouped all the sub-kinds and then performed SVM and K-nearest neighbour. They found the better result in SVM.	Two different types of classification techniques are applied in the field of detecting the white Blood Cells in cancer disease.	Hybrid Model: This study had used SVM and KNN as the classifier in the sub types of ALC and AML.
21	Akbulut et al. [57]	In this study, various classification methods were applied. The different power of alpha and beta frequency bands was used on different types of classification like LDA, NB, KNN, SVM, DT and LR.	This study discussed the pattern recognition by classifying the EEG signals. They did evaluation on the neural interaction during the hand movement present or absent in voluntary mode.	Different types of classification techniques were applied on detecting the presence and absence of hand movement through EEG signals.	Hybrid Model:- The Various types of classifier were used for identifying the best accuracy of result among them.
22	Saha et al. [58]	This paper adopted different machine learning models like K-nearest neighbour, Decision tree, SVM, Discriminant Analysis and logistic regression for performing the efficiency of the extracted heart signals feature.	In detection to the heart signals, ECG Signals were used. These ECG signals were further used to extract features. Various models were used to check the efficiency.	Different classifiers were used in the field of classifying the heart signals.	Hybrid model: The various ML models like DA, DT, LR, SVM and KNN were applied on the ECG signals.
23	Vaishnave et al. [59]	In this paper the authors tried to replace the K nearest neighbour with SVM so that the performance can be increased.	In this study the researchers tried to categorize the diseases present in groundnut leaf. They used the image pre-processing Further	The KNN and SVM were applied for performance improvement in detecting and classifying the disease present in groundnut leaf.	Hybrid Model: The study was applied initially in KNN and after that for improving the performance they applied SVM classifier.

Sr. no.	Authors	Method	Approach	Domain applied	Single/hybrid
			they tried to improve the performance of KNN with the help of SVM.		
24	Preamthaisong et al. [60]	A hybrid classification of logistic regression, self-organizing map, K-nearest neighbour, SVM, decision tree, RF, neural network were used in the study for detecting the distributed denial of service present in software denied network.	They proposed the combination of genetic algorithm with decision tree and found the good result accuracy present in it.	Different types of classifiers were used by the authors in distributed denial of service.	Hybrid Model: This study has used various types of classifiers for identifying the higher accuracy result of with genetic algorithm in combination with the decision tree.
25	Sinayobye et al. [61]	This study has applied seven different classifiers i.e. SVM, decision tree, NB Linear Discriminant analysis, K-nearest neighbour and logistic regression (LR) to identify the best feature selection.	This paper tried to identify the performance of all applied classifiers in search of best feature selection according to the accuracy, precision, specificity, F-measure, sensitivity etc. RF classifier was found good.	All types of classifiers are applied on the dataset of smart meter.	Hybrid Model: The study identified the best performance among seven different types of classifiers.
26	Karim et al. [62]	This study used the unimodal and multi-model features for each type of input and they trained them on various classifiers like NB, logistic regression, RF, gradient boosting tree, SVM, decision tree, K-NN.	They used the new approach on data taken from the cancer Atlas and segregate them on the basis of subtypes and survival rates of patients. They performed training and testing on various classifiers.	Different classifier were applied on the various subtypes of patients suffering from breast cancer.	Hybrid Model: The seven different classifiers were adopted as the baseline models of machine learning.
27.	Askarian et al. [63]	This study used the SVM as the learning technique. They proposed the method in detecting the disease caused due to tooth loss.	They applied a detection method by the use of smart phones where the images were processed and a SVM classifier was applied to distinguish the healthy gum and infected gums.	The different types of classification techniques were applied on tooth loss.	Hybrid Model: The different classification techniques like CNN, SVM and K-nearest neighbour were applied.

Sr. no.	Authors	Method	Approach	Domain applied	Single/hybrid
28	Verma et al. [64]	This study used the K-mean clustering, SVM and conventional neural network method for identifying the infection or disease in the rice crop.	The study used the different classification techniques for identifying the disease in crops. Finally, the various classification techniques were applied on the extracted features.	Different types of classification techniques are used in the identification of disease present in the rice crops.	Hybrid model: SVM, K-means and CNN were applied.
29	Ramya et al.[65]	This paper used the K-means, SVM and neural network as the classifier in detecting the disease in fruit with the help of image processing and cloud computing.	They detected the data with the help of image processing and some data was stored by cloud computing for further usage in the agricultural fields.	Different types of classification techniques are applied in detecting the disease present in fruit.	Hybrid Model: The different types of techniques like K-means, neural network and SVMs were used.
30	Bhatia et al. [66]	In this paper author applied the SVM along with the logistic regression algorithm. They used the SVM for reducing the noise and further that data was fed to Logistic Regression.	In the study the user tried to predict the disease in the tomato plant. For this they took the mild dew powdery dataset and perform random over sampling. Further they applied the various classifiers for noise removal.	The combination of these various algorithms are applied to perform the prediction of disease in tomato plant.	Hybrid: The study applied SVM, K-nearest neighbour, LDA and logistic regression in disease identification of tomato plant.
31	Rahamathunnisa et al. [67]	In this method the authors had applied the K-means technique for image segmentation and SVM for classifying the image.	They extracted the features of vegetables in the form of colour, size and shapes. Further the segmentation process was applied through K-NN and with the help of SVM they classified the images.	The K-nearest neighbour and SVM classifiers were used for identifying the disease present in the vegetables.	Hybrid Model: This study used the K-nearest neighbour and SVM classifier for segmentation and classification.

Table 2 Analysis based on the classification performances

Sr. no.	Reference	Method used	Result	Best method (if any)
1	[68]	KNN SVM DT RF Gradient Boosting	98%	Gradient Boosting Algorithm for Fall Detection
2	[69]	LDA DT SVM KNN NB	99.9%	DT for Mosquito Identification
3	[70]	NB DT RF KNN Gated Recurrent Unit CNN 3 Way Convolutional Recurrent Neural Network	87.75%	NB classifier used for sentiment classification
4	[71]	Different Models of SVM: SVM_model_1 SVM_model_2 SVM_model_3 SVM_model_4 SVM_model_5 SVM_model_6	90.34%	Accuracy in SVM_model_5 for Pancreatic Cancer Detection
5	[72]	SVM LR MLP RF LR NB Their proposed model	98.5%	Their proposed model with feature selection and weighting methods applied for heart disease prediction
6	[73]	NB Logistic Regression SVM Decision Tree Recursive Feature Elimination K-Nearest Neighbour RF	94.45%	Recursive feature elimination in heart disease prediction
7	[74]	SVM KNN Neural Network RF CNN	NA	CNN obtained the best accuracy result
8	[75]	Lasso regression RF AdaBoost SVM Partial Least Square	Good mean value test	Lasso regression best performing algo. for traffic- related pollutants

Sr. no.	Reference	Method used	Result	Best method (if any)
9	[76]	Regularised logistic regression ANN SVM Adaboost Partial Least Squares-Discriminant Analysis	NA	Regularised Logistic Regression provides good result
10	[77]	Used CNN method with three object detection architectures: Faster R-Scan SSD YOLOv3	99.8%	YOLOv3 as a best detection accuracy for diagnosis of Alzheimer's disease
11	[78]	RF SVM LSTM (2 hidden Layers) Bi LSTM (2 hidden Layers) LSTM (4 hidden Layers) Bi LSTM (4 hidden Layers)	99.12%	Data trained at bidirectional LSTM (4 Hidden Layers) has the high accuracy result in Detecting the irregularities in heart rhythm
12	[79]	Four Classifiers used: LR KNN SVM CNN	98%	CNN gives the best result in detecting the disease in plant leaf.
13	[80]	LR KNN DT RF Tree AdaBoost Classifier XGBoost Classifier LightGBM Classifier Multilayer Perceptron	88%	RF after outlier removal gives good result in patients of liver disease
14	[81]	SVM KNN	100%	KNN is preferred on detecting medicinal plants
15	[82]	SVM CNN	98.38%	SVM for identifying the disease in rice leaf
16	[83]	Neural Networks	98.89%	Accuracy obtained in this proposed system in detecting eye disorder
17	[84]	Four models used: CNN VGG16 VGG19 InceptionV3	97%	All models provide accuracy of 97% in detecting the pneumonia
18	[85]	CNN Compared with Traditional Methods: NB RF SVM LR Feature Selection	90%	Encouraging result given by CNN for detecting unexpected findings in the reports of radiology
19	[86]	Principal Component Analysis SVM	86%	Provides SVM in detecting the disease in data tele monitoring

Sr. no.	Reference	Method used	Result	Best method (if any)
20	[87]	Machine learning methods used: ANNs SVM	94.34%	ANNs model with a combination of feature extraction with SURF gives the best result in mildew detection of powdery strawberry
21	[88]	SVM Fully connected Neural Network KNN , CNN	99.3%	CNN provides the good accuracy result in disease detection of image-based plant
22	[89]	ANN SVM LR	97.59%	ANN classifier gives better performance in recognizing the diabetic patients
23	[90]	LR RF CNN	93.3%	CNN worked best in identifying the parkinson's disease
24	[91]	SVM RF KNN Long Short-Term Memory	94.4%	SVM gives the best result the classification of ECG signals
25	[92]	DT RF SVM NB KNN	NA	RF classifier gives the best accuracy result for predicting the depression, stress and anxiety
26	[93]	LDA SVM RF	66.54%	LDA gives better result than other two in the diagnosis of PTSD.

5. Conclusion

In this paper machine learning algorithms have been discussed and analyzed. It has been discussed and analyzed based on different computational aspects. It also includes different domain cases in terms of methodological exploration. This paper elaborates the current scenario with a review and analysis of the latest trends considering machine learning approaches. It elaborates the computational, analytical along with the methodological prospect for the understanding of the role of machine learning and applicability. Based on the analysis, gaps have been analyzed and discussed. The overall analysis indicates that the hybrid model may be better in comparison to all. In terms of methods decision tree, NB classifier, SVM, CNN, RF and KNN are found to be prominent in different application domains.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

- [1] Ray S. A quick review of machine learning algorithms. In international conference on machine learning, big data, cloud and parallel computing 2019 (pp. 35-9). IEEE.
- [2] Ghadge R, Kulkarni J, More P, Nene S, Priya RL. Prediction of crop yield using machine learning. International Research Journal of Engineering and Technology. 2018; 5(2):2237-9.
- [3] Cheng S, Liu J, Tang X. Using unlabeled data to improve inductive models by incorporating transductive models. Editorial Preface. 2014; 3(2):33-8.
- [4] Kishore R, Kaur T. Backpropagation algorithm: an artificial neural network approach for pattern recognition. International Journal of Scientific & Engineering Research. 2012; 3(6):6-9.
- [5] Mishra A, Mohapatro M. An IoT framework for bio-medical sensor data acquisition and machine learning for early detection. International Journal of Advanced Technology and Engineering Exploration. 2019; 6(54):112-25.
- [6] Farsad N, Goldsmith A. Neural network detection of data sequences in communication systems. IEEE Transactions on Signal Processing. 2018; 66(21):5663-78.

- [7] Dubey AK, Kumar A, Agrawal R. An efficient ACO-PSO-based framework for data classification and preprocessing in big data. *Evolutionary Intelligence*. 2020;1-4.
- [8] Sidey-Gibbons JA, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Medical Research Methodology*. 2019.
- [9] Sebastiani F. Machine learning in automated text categorization. *ACM Computing Surveys (CSUR)*. 2002; 34(1):1-47.
- [10] <https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi>. Accessed 28 May 2020.
- [11] Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. *Radiographics*. 2017; 37(2):505-15.
- [12] Patel E, Chauhan S. Raag detection in music using supervised machine learning approach. *International Journal of Advanced Technology and Engineering Exploration*. 2017; 4(29):58-67.
- [13] Shokri R, Stronati M, Song C, Shmatikov V. Membership inference attacks against machine learning models. In *symposium on security and privacy 2017* (pp. 3-18). IEEE.
- [14] Camacho DM, Collins KM, Powers RK, Costello JC, Collins JJ. Next-generation machine learning for biological networks. *Cell*. 2018; 173(7):1581-92.
- [15] Hussin SK, Omar YM, Abdelmageid SM, Marie MI. Traditional machine learning and big data analytics in virtual screening: a comparative study. *International Journal of Advanced Computer Research*. 2020; 10(47):72-88.
- [16] Sun S. A survey of multi-view machine learning. *Neural Computing and Applications*. 2013; 23(7-8):2031-8.
- [17] Mduma N, Kalegele K, Machuve D. Machine learning approach for reducing students dropout rates. *International Journal of Advanced Computer Research*. 2019; 9(42): 156-69.
- [18] Mathur S, Badone A. A methodological study and analysis of machine learning algorithms. *International Journal of Advanced Technology and Engineering Exploration*. 2019; 6(51):45-9.
- [19] Postnikov EB, Esmedjaeva DA, Lavrova AI. A CatBoost machine learning for prognosis of pathogen's drug resistance in pulmonary tuberculosis. In *global conference on life sciences and technologies 2020* (pp. 86-7). IEEE.
- [20] Panigrahi KP, Sahoo AK, Das H. A CNN approach for corn leaves disease detection to support digital agricultural system. In *international conference on trends in electronics and informatics 2020* (pp. 678-83). IEEE.
- [21] Jamil S, Abbas MS, Umair M, Habib F, Hussain Y. A novel deep neural network CanNet for malignant detection. In *international conference on information science and communication technology 2020* (pp. 1-5). IEEE.
- [22] Aruchamy S, Haridasan A, Verma A, Bhattacharjee P, Nandy SN, Vadali SR. Alzheimer's disease detection using machine learning techniques in 3D MR images. In *national conference on emerging trends on sustainable technology and engineering applications 2020* (pp. 1-4). IEEE.
- [23] Iadanza E, Goretti F, Sorelli M, Melillo P, Pecchia L, Simonelli F, et al. Automatic detection of genetic diseases in pediatric age using pupillometry. *IEEE Access*. 2020; 8:34949-61.
- [24] Mane H, Ghorpade P, Bahel V. Computational intelligence based model detection of disease using chest radiographs. In *international conference on emerging trends in information technology and engineering 2020* (pp. 1-5). IEEE.
- [25] Roopa NK, Mamatha GS. Correlating the machine learning models for automatic error detection and correction in medical Images. In *international conference on innovative mechanisms for industry applications 2020* (pp. 64-70). IEEE.
- [26] Majdi MS, Salman KN, Morris MF, Merchant NC, Rodriguez JJ. Deep learning classification of chest x-ray images. In *southwest symposium on image analysis and interpretation 2020* (pp. 116-9). IEEE.
- [27] Ramírez VM, Kmetzsch V, Forbes F, Dojat M. Deep learning models to study the early stages of parkinson's disease. In *international symposium on biomedical imaging 2020* (pp. 1534-7). IEEE.
- [28] Watanabe M, Sugiura Y, Saito H, Koyama T, Fujita K. Detection of cervical myelopathy with Leap Motion Sensor by random forests. In *global conference on life sciences and technologies 2020* (pp. 214-6). IEEE.
- [29] Ramos-Giraldo P, Reberg-Horton C, Locke AM, Mirsky S, Lobaton E. Drought stress detection using low-cost computer vision systems and machine learning techniques. *IT Professional*. 2020; 22(3):27-9.
- [30] Ratley A, Minj J, Patre P. Leukemia disease detection and classification using machine learning approaches: a review. In *first international conference on power, control and computing technologies 2020* (pp. 161-5). IEEE.
- [31] Kundu N, Rani G, Dhaka VS. Machine learning and IoT based disease predictor and alert generator system. In *fourth international conference on computing methodologies and communication 2020* (pp. 764-9). IEEE.
- [32] Bini D, Pamela D, Prince S. Machine vision and machine learning for intelligent Agrobots: a review. In *international conference on devices, circuits and systems 2020* (pp. 12-16). IEEE.
- [33] Kundu AK, Fattah SA, Wahid KA. Multiple linear discriminant models for extracting salient characteristic patterns in capsule endoscopy images for multi-disease detection. *IEEE Journal of Translational Engineering in Health and Medicine*. 2020; 8:1-1.
- [34] Nidigattu GR, Mattela G, Jana S. Non-invasive modeling of heart rate and blood pressure from a photoplethysmography by using machine learning techniques. In *international conference on communication systems & networks 2020* (pp. 7-12). IEEE.

- [35] Tuncer T, Dogan S, Özyurt F, Belhaouari SB, Bensmail H. Novel multi center and threshold ternary pattern based method for disease detection method using voice. *IEEE Access*. 2020; 8:84532-40.
- [36] Iqbal J, Iqbal MM, Khadam U, Nawaz A. Ordinary learning method for heart disease detection using clinical data. In international conference on computing, mathematics and engineering technologies 2020 (pp. 1-6). IEEE.
- [37] Kumari RS, Jose JP. Seizure detection in EEG using time frequency analysis and SVM. In international conference on emerging trends in electrical and computer technology 2011 (pp. 626-30). IEEE.
- [38] Shahbakhti M, Taherifar D, Zareei Z. Combination of PCA and SVM for diagnosis of parkinson's disease. In international conference on advances in biomedical engineering 2013 (pp. 137-40). IEEE.
- [39] Bourouhou A, Jilbab A, Nacir C, Hammouch A. Comparison of classification methods to detect the Parkinson disease. In international conference on electrical and information technologies 2016 (pp. 421-4). IEEE.
- [40] Padol PB, Sawant SD. Fusion classification technique used to detect downy and Powdery Mildew grape leaf diseases. In international conference on global trends in signal processing, information computing and communication 2016 (pp. 298-301). IEEE.
- [41] Shetty S, Rao YS. SVM based machine learning approach to identify Parkinson's disease using gait analysis. In international conference on inventive computation technologies 2016 (pp. 1-5). IEEE.
- [42] Padol PB, Yadav AA. SVM classifier based grape leaf disease detection. In conference on advances in signal processing 2016 (pp. 175-9). IEEE.
- [43] Patil P, Yaligar N, Meena SM. Comparison of performance of classifiers-SVM, RF and ANN in potato blight disease detection using leaf images. In international conference on computational intelligence and computing research 2017 (pp. 1-5). IEEE.
- [44] Prakash RM, Saraswathy GP, Ramalakshmi G, Mangaleswari KH, Kaviya T. Detection of leaf diseases and classification using digital image processing. In international conference on innovations in information, embedded and communication systems 2017 (pp. 1-4). IEEE.
- [45] Islam M, Dinh A, Wahid K, Bhowmik P. Detection of potato diseases using image segmentation and multiclass support vector machine. In Canadian conference on electrical and computer engineering 2017 (pp. 1-4). IEEE.
- [46] Masazhar AN, Kamal MM. Digital image processing technique for palm oil leaf disease detection using multiclass SVM classifier. In international conference on smart instrumentation, measurement and application 2017 (pp. 1-6). IEEE.
- [47] Mathew AR, Anto PB. Tumor detection and classification of MRI brain image using wavelet transform and SVM. In international conference on signal processing and communication 2017 (pp. 75-8). IEEE.
- [48] Karan B, Mahto K, Sahu SS. Detection of parkinson disease using variational mode decomposition of speech signal. In international conference on communication and signal processing 2018 (pp. 508-12). IEEE.
- [49] Kiss G, Takács AB, Sztahó D, Vicsi K. Detection possibilities of depression and Parkinson's disease based on the ratio of transient parts of the speech. In international conference on cognitive info communications 2018 (pp. 000165-8). IEEE.
- [50] Bhimte NR, Thool VR. Diseases detection of cotton leaf spot using image processing and SVM classifier. In second international conference on intelligent computing and control systems 2018 (pp. 340-4). IEEE.
- [51] Kusumo BS, Heryana A, Mahendra O, Pardede HF. Machine learning-based for automatic detection of corn-plant diseases using image processing. In international conference on computer, control, informatics and its applications 2018 (pp. 93-7). IEEE.
- [52] Win KY, Maneerat N, Choomchuay S, Sreng S, HAMAMOTO K. Suitable supervised machine learning techniques for malignant mesothelioma diagnosis. In biomedical engineering international conference 2018 (pp. 1-5). IEEE.
- [53] Pardede HF, Suryawati E, Sustika R, Zilvan V. Unsupervised convolutional autoencoder-based feature learning for automatic detection of plant diseases. In international conference on computer, control, informatics and its applications 2018 (pp. 158-62). IEEE.
- [54] Gurralla KK, Yemineni L, Rayana KS, Vajja LK. A new segmentation method for plant disease diagnosis. In international conference on intelligent communication and computational techniques 2019 (pp. 137-41). IEEE.
- [55] Al Qadi L, El Rifai H, Obaid S, Elnagar A. Arabic text classification of news articles using classical supervised classifiers. In international conference on new trends in computing sciences 2019 (pp. 1-6). IEEE.
- [56] Lakshmi KG, Naik NM. Automated detection of white blood cells cancer disease. In international conference on advances in information technology 2019 (pp. 24-8). IEEE.
- [57] Akbulut H, Güney S, Çotuk HB, Duru AD. Classification of EEG signals using alpha and beta frequency power during voluntary hand movement. In scientific meeting on electrical-electronics & biomedical engineering and computer science 2019 (pp. 1-4). IEEE.
- [58] Saha R, Singh RK, Kumar R, Singh G, Goel T, Pal PK. Classification of human heart signals by novel feature extraction techniques for rescue application. In fifth international conference on image information processing 2019 (pp. 156-60). IEEE.
- [59] Vaishnave MP, Devi KS, Srinivasan P, ArutPerumJothi G. Detection and Classification of Groundnut Leaf Diseases using KNN classifier. In

- international conference on system, computation, automation and networking 2019 (pp. 1-5). IEEE.
- [60] Preamthaisong P, Auyporntrakool A, Aimtongkham P, Sriwuttisap T, So-In C. Enhanced DDoS detection using hybrid genetic algorithm and decision tree for SDN. In 16th international joint conference on computer science and software engineering 2019 (pp. 152-7). IEEE.
- [61] Sinayobye JO, Kyanda SK, Kiwanuka NF, Musabe R. Hybrid model of correlation based filter feature selection and machine learning classifiers applied on smart meter data set. In symposium on software engineering in Africa 2019 (pp. 1-10). IEEE.
- [62] Karim MR, Wicaksono G, Costa IG, Decker S, Beyan O. Prognostically relevant subtypes and survival prediction for breast cancer based on multimodal genomics data. *IEEE Access*. 2019; 7:133850-64.
- [63] Askarian B, Tabei F, Tipton GA, Chong JW. Smartphone-based method for detecting periodontal disease. In healthcare innovations and point of care technologies 2019 (pp. 53-5). IEEE.
- [64] Verma G, Taluja C, Saxena AK. Vision based detection and classification of disease on rice crops using convolutional neural network. In international conference on cutting-edge technologies in engineering 2019 (pp. 1-4). IEEE.
- [65] Ramya R, Kumar P, Sivanandam K, Babykala M. Detection and classification of fruit diseases using image processing & cloud computing. In international conference on computer communication and informatics 2020 (pp. 1-6). IEEE.
- [66] Bhatia A, Chug A, Singh AP. Hybrid SVM-LR Classifier for powdery mildew disease prediction in tomato plant. In international conference on signal processing and integrated networks 2020 (pp. 218-23). IEEE.
- [67] Rahamathunnisa U, Nallakaruppan MK, Anith A, Kumar KS. Vegetable disease detection using k-means clustering and SVM. In international conference on advanced computing and communication systems 2020 (pp. 1308-11). IEEE.
- [68] Zurbuchen N, Bruegger P, Wilde A. A comparison of machine learning algorithms for fall detection using wearable sensors. In international conference on artificial intelligence in information and communication 2020 (pp. 427-31). IEEE.
- [69] Genoud AP, Gao Y, Williams GM, Thomas BP. A comparison of supervised machine learning algorithms for mosquito identification from backscattered optical signals. *Ecological Informatics*. 2020.
- [70] Basiri ME, Abdar M, Cifci MA, Nemati S, Acharya UR. A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques. *Knowledge-Based Systems*. 2020.
- [71] Briones E, Lao A, Solano GA. A Pancreatic cancer detection support tool using mass spectrometry data and support vector machines. In international conference on artificial intelligence and signal processing 2020 (pp. 1-6). IEEE.
- [72] Ali F, El-Sappagh S, Islam SR, Kwak D, Ali A, Imran M, et al. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. *Information Fusion*. 2020; 63:208-22.
- [73] Kaur PC. A study on role of machine learning in detecting heart disease. In fourth international conference on computing methodologies and communication 2020 (pp. 188-93). IEEE.
- [74] Murat F, Yildirim O, Talo M, Baloglu UB, Demir Y, Acharya UR. Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review. *Computers in Biology and Medicine*. 2020.
- [75] Šimić I, Lovrić M, Godec R, Kröll M, Bešlić I. Applying machine learning methods to better understand, model and estimate mass concentrations of traffic-related pollutants at a typical street canyon. *Environmental Pollution*. 2020.
- [76] Matta K, Vigneau E, Cariou V, Mouret D, Ploteau S, Le Bizec B, et al. Associations between persistent organic pollutants and endometriosis: a multipollutant assessment using machine learning algorithms. *Environmental Pollution*. 2020.
- [77] Fong JX, Shapiai MI, Tiew YY, Batool U, Fauzi H. Bypassing MRI pre-processing in alzheimer's disease diagnosis using deep learning detection network. In international colloquium on signal processing & its applications 2020 (pp. 219-24). IEEE.
- [78] Singh S, Penzel T. Classification and detection of heart rhythm irregularities using machine learning. In first international conference on power, control and computing technologies 2020 (pp. 438-42). IEEE.
- [79] Sharma P, Hans P, Gupta SC. Classification of plant leaf diseases using machine learning and image preprocessing techniques. In international conference on cloud computing, data science & engineering 2020 (pp. 480-4). IEEE.
- [80] Kuzhippallil MA, Joseph C, Kannan A. Comparative analysis of machine learning techniques for indian liver disease patients. In international conference on advanced computing and communication systems 2020 (pp. 778-82). IEEE.
- [81] Raghukumar AM, Narayanan G. Comparison of machine learning algorithms for detection of medicinal plants. In fourth international conference on computing methodologies and communication 2020 (pp. 56-60). IEEE.
- [82] Sethy PK, Barpanda NK, Rath AK, Behera SK. Deep feature based rice leaf disease identification using support vector machine. *Computers and Electronics in Agriculture*. 2020.
- [83] Deepa R, Narayanan NK. Detection of microaneurysm in retina image using machine learning approach. In international conference on innovative trends in information technology 2020 (pp. 1-5). IEEE.
- [84] Labhane G, Pansare R, Maheshwari S, Tiwari R, Shukla A. Detection of pediatric pneumonia from chest X-Ray images using CNN and transfer learning. In international conference on emerging technologies

- in computer engineering: machine learning and internet of things 2020 (pp. 85-92). IEEE.
- [85] López-Úbeda P, Díaz-Galiano MC, Martín-Noguerol T, Ureña-López A, Martín-Valdivia MT, Luna A. Detection of unexpected findings in radiology reports: a comparative study of machine learning approaches. *Expert Systems with Applications*. 2020; 160:113647.
- [86] Kobayashi N, Ishikawa M, Okazaki H, Homma S. Disease detection using machine learning in vital sign data telemonitoring. In 2nd global conference on life sciences and technologies 2020 (pp. 309-10). IEEE.
- [87] Shin J, Chang YK, Heung B, Nguyen-Quang T, Price GW, Al-Mallahi A. Effect of directional augmentation using supervised machine learning technologies: a case study of strawberry powdery mildew detection. *Biosystems Engineering*. 2020; 194:49-60.
- [88] Radovanović D, Đukanović S. Image-based plant disease detection: a comparison of deep learning and classical machine learning algorithms. In international conference on information technology 2020 (pp. 1-4). IEEE.
- [89] Chakour I, El Mourabit Y, Daoui C, Baslam M. Multi-agent system based on machine learning for early diagnosis of diabetes. In international conference on optimization and applications 2020 (pp. 1-6). IEEE.
- [90] Chakraborty S, Aich S, Han E, Park J, Kim HC. Parkinson's disease detection from spiral and wave drawings using convolutional neural networks: a multistage classifier approach. In international conference on advanced communication technology 2020 (pp. 298-303). IEEE.
- [91] Pandey SK, Janghel RR, Vani V. Patient specific machine learning models for ECG signal classification. *Procedia Computer Science*. 2020; 167:2181-90.

- [92] Priya A, Garg S, Tigga NP. Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Computer Science*. 2020; 167:1258-67.
- [93] Kim YW, Kim S, Shim M, Jin MJ, Jeon H, Lee SH, Im CH. Riemannian classifier enhances the accuracy of machine-learning-based diagnosis of PTSD using resting EEG. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. 2020.



Ravita Chahar is an Assistant Professor in Computer Science and Engineering Department at Chitkara University, Punjab. She is having a rich experience of 12 years in the teaching field. She holds the degree of Master in Engineering in Computer Science and Engineering from Maharshi Dayanand University, Rohtak. Her area of research interests are Databases, Data Mining and Machine Learning. Currently, she is pursuing Ph.D in Computer Science and Engineering from the Chitkara University, Punjab.
Email: ravita.chahar@chitkara.edu.in



Deepinder Kaur is an Assistant Professor in Computer Science and Engineering Department at Chandigarh University, Punjab. She is having 11 years experience in teaching field. She holds the degree of Master in Technology in Computer Science and Engineering from Punjab Technical University, Jalandhar. Her area of research interests are Databases, Data Mining and Software Engineering.