IDENTIFICATION OF PARAMETERS FOR CLASSIFICATION OF COVID-19 PATIENT'S RECOVERY DAYS USING MACHINE LEARNING TECHNIQUES

DIGAMBAR UPHADE[1], ANIKET MULEY[2,*]

[1]Department of Statistics, KRT Arts, BH Commerce and AM Science College, Nashik-422002 (M.S.) India

[2]School of Mathematical Sciences, Swami Ramanand Teerth Marathwada University, Nanded-431606 (M.S.) India

Abstract: Nowadays, Corona virus has been spreading all over the world. Discovery of various perspectives is going on. Our aim is to identify the recovery days of patients from the Covid-19 disease. To classify the patient using various parameters that affect his/her recovery days. It is complex to deal with numerous parameters, so to reduce the complexity feature selection techniques were employed. In this study, we have dealt with different machine learning approaches for classifying the patients dataset collected through the online survey methodology. We are pioneers in dealing with aspects. Based on these techniques, our interest is to classify the patients as based on the number of recovery days. This present study has major contributions as a method of classification and is an easily understandable way using statistical visualization plots viz., bar plots, pie charts etc. The machine learning algorithms like Logistic regression, Decision tree, Random forest, Neural network, Support vector machine, K Nearest Neighbor were used for performing this task. Further, comparative study is performed and the neural network gives better accuracy to classify the respondents. Finally, results explored with supervised learning are more accurate to detect the COVID-19 recovery patients' cases and neural network is found to be an efficient algorithm as compared with other algorithms (100%).

*Corresponding author

E-mail address: aniket.muley@gmail.com

1. INTRODUCTION

Nowadays, Corona caused things that never happened in the world. To overcome the pandemic situation there is still no exact treatment to recover and how to prevent corona disease. It is a standard practice to recognize the symptoms of the corona and in general, how you would become treated to be cured from it after a certain number of days. There is an essential aspect of this understanding that, after how many days the patient is said to recover after he or she is found to be positive. Also, it is important that how much cost it will require for overall treatment? Some researchers have worked on the similar aspect but in a different view, as per our perception we are the first who have focused on Covid-19 patient's recovery day's prediction perspective.

Malla and Alphonse [23] have proposed a majority voting technique based Ensemble Deep Learning (MVEDL) model to identify informative tweets for COVID-19 dataset. Khanday et al. [1] applied machine learning techniques for detection of COVID-19 classification obtained through textual clinical reports. They found that Logistic regression and Multinomial Naive Bayes gives better results as compared to other Machine learning algorithms. Rogier et al. [25] have performed comparative study characteristics of suspected hospitalized COVID-19 patients. COVID-19 negative patients were observed more often than active smokers and had less often cough, fever and digestive symptoms as compared to the positive ones. COVID-19 positive patients had higher median Neutrophil and Lymphocyte counts and lower CRP level. Garrido et al. [18] observed that the COVID-19 pandemic has negatively affected the professional and home lives of oncologists, especially in women. Reduced research time for female oncologists may have long-lasting career consequences, especially for those at key stages in their career. Further, gender gap for promotion to leadership positions may widen further as a result of the pandemic has been observed. Khaneshpour et al. [13] highlighted the transmission of virus, its clinical signs, laboratory characteristics, the pathogenicity of it, vaccines and apart from that the prevention and control of the spread of the virus.

Shorten et al. [6] presented an application of Deep Learning to fight COVID-19. SARS-CoV-2 and COVID-19 have brought about many new problems for humanity to solve. Munir et al. [17] discussed different methods used for the detection of COVID-19; in addition, they proposed a deep neural network for it. Karaaslan and Aydin [10] studied limitations and challenges on the diagnosis of COVID-19 using radiology images and deep learning. Gupta et al. [19] analyzed the outbreak of COVID-19 disease and trained for Indian region and testing has been done for the number of cases for the next three weeks. SEIR and regression machine learning models were used for predictions and the performance was evaluated using RMSLE. Quintero et al. [19] discussed approaches associated with the development of predictive models for the SEIRD variables based on the historical data. The SEIRD predictive models encompass a deep analysis of the dependence of the variables. Claudio et al. [5] elaborates the possibility of improving results related to COVID-19 chest X-ray image identification with deep learning-based approaches. Ogundokun et al. [20] proposed a simple average aggregated machine learning method to predict the number, size and length of COVID-19 cases extent and wind-up period crosswise India. They examined the data through the Autoregressive Integrated Moving Average Model (ARIMA) and built a simple mean aggregated method established on the performance of 3 regression techniques viz. Support Vector Regression (SVR), Artificial Neural Network (ANN) and Linear Regression (LR). Pyrros et al. [2] used Deep learning analysis of single frontal chest radiographs to generate combined co-morbidity and pneumonia scores that predict the need for supplemental oxygen and hospitalization with COVID-19 infection. Gois et al. [12] focused on the survey of epidemic forecasts associated with the prediction of COVID-19 statistics viz., several infections and deaths, spread locations and others with the preliminaries of SEIR, SIR, Facebook Prophet, Kalman Filters (KFs) and long short term memory (LSTM) models used by SESA to COVID-19. Bekele et al. [11] observed the overall high psychological impact of COVID-19 pandemic among healthcare workers, community and patients. Gambhir et al. [9] analyzed the current trend of the transmission of Covid-19 in the world. A comprehensive study of the spread of the virus outbreak situation in India is performed with a polynomial regression and support vector machine and it gives more accuracy. Ramanathan

and Ramasundaram [24] performed computation of COVID-19 rRT-PCR positive test dataset via classification through textual big data mining with machine learning. The machine learning techniques are used to classify the patients, who are tested positive for corona, into three different classes: mild, moderate and severe based on the clinical report of dataset.

Rajagopal [21] proposed work suggests the possibility of using X-ray images of persons having COVID-19 symptoms classified as: healthy, COVID-19 affected, or pneumonia affected. Classification is carried out using Convolutional Neural Network (CNN), transfer learning using VGG Net and machine learning techniques viz., SVM and XGBoost which utilizes features extracted with the help of Convolutional Neural Network. Thimmegowda and Kumar [4] comprehensively reviewed various Covid-19 detection techniques using ML and DL. The Machine Learning approaches employed various models like RF, ARIMA, SVR, CUBIST and Gradient Boosting to precisely make predictions. Rasheed et al. [16] investigated the potential of machine learning methods for automatic diagnosis of corona virus with high accuracy from X-ray images. In their proposed CNN with PCA technique eclipsed the ultra-modern and advanced approaches as it attained highest accuracy. Kwekha-Rashid et al. [3] studied the role of machine-learning algorithms to optimize COVID-19 studies that had been published during 2020 and were related to this topic by seeking in Science Direct, Springer, Hindawi and MDPI using machine learning, supervised learning and unsupervised learning as keywords. Yasar and Ceylan [14] performed a comparative study for detection of Covid-19 on CT lung images using texture analysis, machine learning and deep learning methods. In their study, CNN was used for classification of CT images of lungs for early diagnosis of Covid-19 disease. Along with that, k-Nearest Neighbors (K-NN) and SVM were used to compare the classification. Ayyildiz et al. [8] observed that, Artificial Neural Network (ANN) is an effective approach to forecast the daily number of people recovered from COVID-19. Thereafter, they have compared with the Autoregressive Integrated Moving Average (ARIMA), Long Short-term Memory (LSTM) and Nonlinear Autoregressive Network with Exogenous Inputs (NARX) methods. Result of their study reveals that Multi-Layer Perceptron (MLP) model gives better forecasting performance for the

number of people recovered from COVID-19. Yadav [22] have utilized six regression analysis based models namely quadratic, third, fourth, fifth and sixth degree and exponential polynomial respectively for the COVID-2019 dataset. They found that the sixth degree polynomial regression models will help Indian doctors and the Government in preparing their plans in the next 7 days and suggested that this model can be tuned for forecasting over long term intervals. Khanday et al. [1] used supervised machine learning techniques for classifying the text into four different categories COVID, SARS, ARDS and both (COVID, ARDS). Further, ensemble learning techniques for classification is also employed. Chaurasia and Pal [26] carried out a study on corona virus patients' number of cases, deaths and recovery cases worldwide within a specific time period of 5 months to predict the future spread of this infectious disease in human society. They have collected a dataset from WHO Corona virus Covid-19 cases and deaths-WHO-COVID19-global-data. Assaf et al. [7] among patients with COVID-19, the ability to identify patients at risk for deterioration during their hospital stay is essential for effective patient allocation and management. Three different machine-learning models were used to predict patient deterioration and compared to predictors and risk prediction score. Malki et al. [28] considered machine learning approaches to predict the spread of the COVID-19 in many countries. A machine learning model has been developed to predict the estimation of the spread of the COVID-19 infection in many countries and the expected period after which the virus can be stopped.

The aim of the study is to identify important features and the best classifier model for the number of recovery days of COVID-19 patient. To achieve this objective we have used a comparative approach of existing machine learning techniques. In the subsequent sections methodology, result and description, and conclusions are discussed in detail.

2. PRELIMINARIES

Here, we have performed a survey based study to identify the Covid-19 patients' recovery days. We had made a questionnaire that included 51 questions related to Covid-19 symptoms and diet. Machine learning is an automated way to perform data analysis in different sectors viz., medical,

engineering, finance, business, education etc. It emerged from artificial intelligence that taught machines from training the datasets. With machine learning we can classify the patterns, evaluate the data and it will help to make decisions with no or least human being interference. Machine learning is broadly classified into three categories: supervised, unsupervised and reinforcement learning. It is observed that, in supervised learning, machines learn via training data. Whereas unsupervised learning algorithms learn with the partitioning or clustering techniques. Reinforcement learning is the mixture of both supervised as well as unsupervised learning algorithms. To analyze recovery day's parameters we have applied ML algorithms and evaluated the accuracy based on featured significant parameters.

2.1 Proposed optimization algorithm:

To perform overall evaluation we have proposed algorithm and the steps are given below:

Step 1: To design a survey questionnaire with structured and unstructured form of questions.

Step 2: To collect the data through online mode with the help of Google form platform.

Step 3: To stop the process of collection of data by considering the appropriate sample size by

Eq. (1): [Ref: 15]

$$n = \frac{z^2 \times \hat{p}(1-\hat{p})}{\varepsilon^2} \tag{1}$$

where, *z -The z score, ε - The margin of error, n - The sample size,*

$\hat{p}$- The population proportion

Step 4: To clean and pre-process the collected data set.

Step 5: To evaluate the data set with unsupervised learning methods viz. graphical representation
i.e. graphs and charts.

Step 6: To identify the significant parameters using features selection techniques viz. number of
the top non-negative features to select we have used Chi-square, Extremely Randomized
tree classifier an extra tree classifier, which will be beneficial to reduce the complexity of
the problem.

Step 7: To check whether the significant parameters are based on pair of arrays (scores, p values) p
value (score > 5 and Total information gain Gini index > 0.025 only). There are three
different situations occurred that are:

Step 7.1: If obtained are the same then goto the next step.

Step 7.2: If obtained parameter counts is different then use a comparative approach with ML algorithms for obtained parameters given the next step.

Step 7.3: If obtained extracted parameters are different in step (7.1-7.2). Combine all the features in a set and then try to apply classifiers mentioned in the next step.

Step 8: To apply various classifier methods viz. Decision Tree, Random Forest, K-NN, Logistic Regression, ANN, Support Vector Machine with significant parameters obtained in step 7.

- **Logistic Regression**: It is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 or 0 binary event utilizing a logit function.[1]

- **Decision Tree:** It is a supervised learning used for both classification and Regression problems. In a decision tree, there are two nodes, which are the decision nodes used to make any decision and have multiple branches, whereas leaf nodes are the output of those decisions and do not contain any further branches. [1]

- **Random Forest**: It is a supervised learning used for both Classification and Regression problems in ML. It is based on ensemble learning i.e. a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. The random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. [1]

- **ANN**: It is derived from Biological neural networks that develop the structure of a human brain. Similar to the human brain that has neurons interconnected to one another. These neurons are known as nodes. It consists of three layers i.e. input Layer, hidden Layer and output Layer. The input goes through a series of transformations using the hidden layer, which finally results in output that is conveyed using this layer.[14, 20]

- **SVM**: It is a Supervised Learning algorithm, which is used for Classification as well as Regression problems. SVM chooses the extreme points/vectors that help in creating the hyperplane and are called as support vectors. [14, 20]

- **K-NN:** It falls under the Supervised Learning category and is used for classification and regression. It is a versatile algorithm also used for imputing missing values and resampling datasets. K-NN to predict the class or continuous value for the new data point. [14]

Step 9: The efficiency of the proposed method is evaluated by means of computing certain performance measures using the following Eq. (2-4):

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \qquad (2)$$

$$Precision = \frac{TP}{(TP + FP)} \qquad (3)$$

$$Recall = \frac{TP}{(TP + FN)} \qquad (4)$$

where, *TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative*.

Step 10: Compare the efficiency of the model to optimize the process.

Step 11: To stop the process.

3. MAIN RESULTS

In this study, primary data is collected through online mode. According to Steps 1-3 in the proposed algorithm the sample size is 171 is considered with the confidence level of 95% (z = 1.96) and that the real value is within ±7.5% of the measured or surveyed value. The basic information obtained as explored through visualization techniques represented in Fig. 1-11.
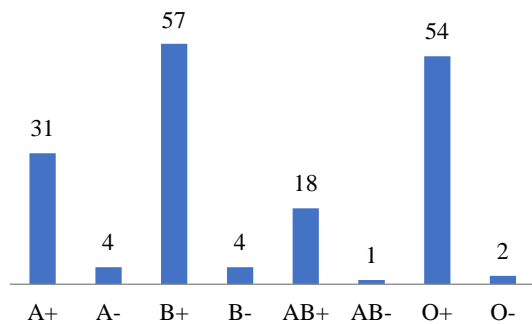

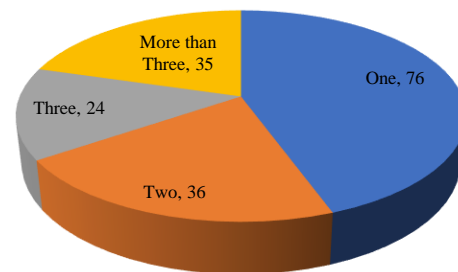Fig. 1 Respondents Blood group classification

Fig. 2 Respondents no. of family members infected
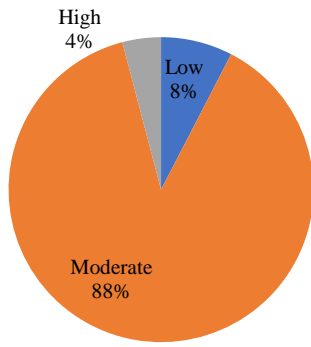
CLASSIFICATION OF PATIENT'S RECOVERY DAYS

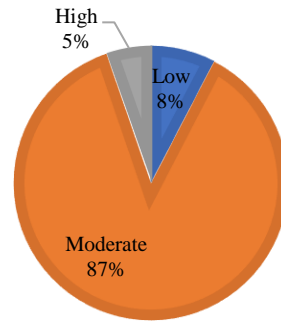Fig. 3 Respondents Blood pressure classification

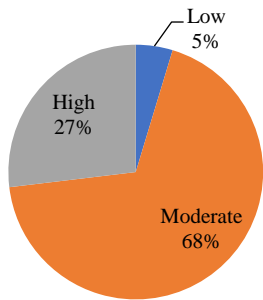Fig. 4 Respondents Blood sugar level classification

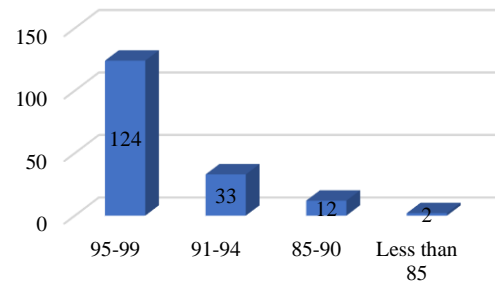Fig. 5 Respondents Body temperature classification

Fig. 6 Respondents Blood oxygen level classification
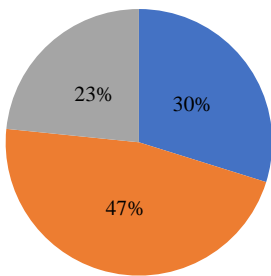
Fig. 7 Respondents diagnosis test for identification
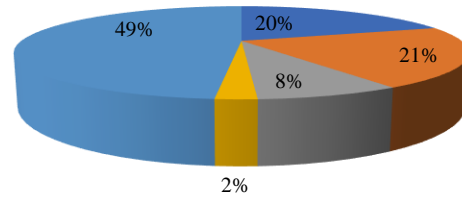
Fig. 8 Respondents HRCT Score test

DIGAMBAR UPHADE, ANIKET MULEY
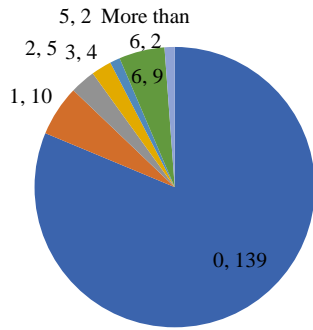
Fig. 9 Respondents taken Remdesivir doses
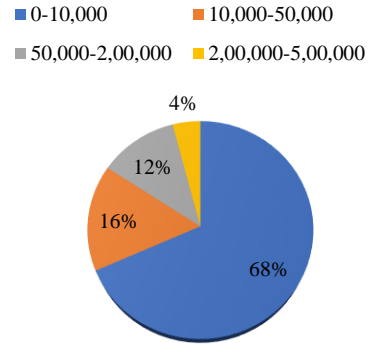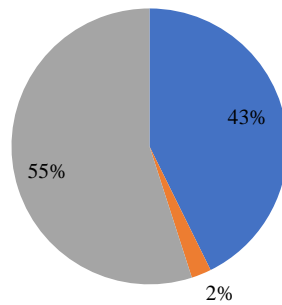
Fig. 10 Respondents overall Treatment Cost

Fig. 11 Respondents classification regarding Ayurvedic medicines
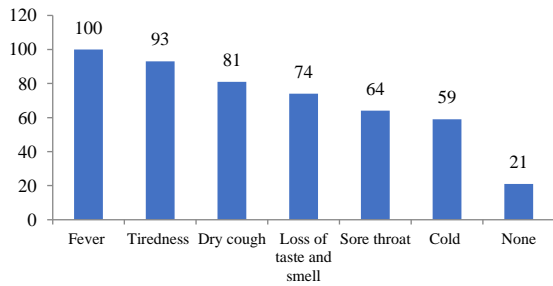
Fig. 12 Respondent observed symptoms
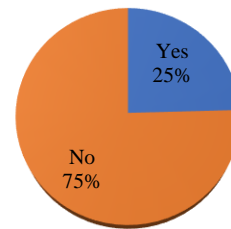
Fig. 13 Requirement of hospitalization

Fig. 1 explains the respondents' blood group classifications and data reveals that B+, O+ and A+ blood group respondent are more prone to have COVID 19 infections. Fig. 2 illustrates that in the majority of the cases it is observed that single family members were infected. Fig. 3 explained that 88% of respondents' blood pressure was moderate during the infected period. Fig. 4 explores that 87 % of the respondents' sugar level was moderate during the infected period. Fig. 5 represents body temperature during the infected period and it is observed that 68 % were moderate and 27% of the respondents had high body temperature. Fig. 6 reveals the information regarding to the data that 87 % of the individuals have blood oxygen levels in the range 95-99 during the infection period. Fig. 7 represents the diversity to diagnose COVID infection and it is observed that 47 % of people prefer the rRT-PCR test, 30% of them would prefer Rapid antigen test and 23 % of the respondents would prefer both of the diagnostic check methods. Fig. 8 explains the HRCT scores of the respondents and explores that 49 % of the cases HRCT were not performed. Further, 0 score observed the 20 % of the respondents having HRCT score between the ranges of 1-7. Also, 21 % of them have HRCT scores between 1-7 and 8 % and subsequent 2 % of the respondents' HRCT score lies between 8-14 and 15-21. Fig. 9 explores the respondents' taken number of Remdesivir doses: 139 out of 171 respondents haven't taken Remdesivir doses and subsequently 1 to 6 doses taken by respondents are least in numbers. Fig.10 explores the amount spent on the treatment by the respondents and 68% of people spent less than Rs. 10,000; 16 % of the respondents spent Rs. 10,000-50,000; 12 % of them spent Rs. 50,000-2,00,000 and 4% of the patients spent Rs. 2 to 5 lacs. Fig. 11 explores the information regarding Ayurvedic medicine taken by the patients. Data reveals that 43% of the respondents had taken Ayurvedic medicine and got beneficial results; 55% of respondents had taken but they didn't get any good results and 2 % of the respondent hasn't taken Ayurvedic medicine. Fig. 12 explores the patients having symptoms from which they are visited to the doctor for further treatment. Fig. 12 reveals that, in our study there are 100 of the respondents having fever, 93 of them having tiredness, 81 of them having dry cough, 74 of them having loss of taste and smell, 64 of them caught sore throat, 59 have coldness problem and 21 does not   have any symptoms but they found to be Covid-19 positive. Fig. 13 explores that, 75 %

of respondents in our study do not require hospitalization whereas 25 % of them required hospitalization.

Table 1 Significant features of the study (Chi-square method)

Feature	Score	P Value
Remdesivir doses	16.843353	0.000041
Treatment costs	9.322834	0.002263
Neutrophil- Lymphocyte Ratio	6.580871	0.010308
Number of days hospitalized	5.711705	0.016852

It is essential to identify the significant features among overall parameters. Table 1 explains the significant features obtained through Chi-square test procedure and it is observed that, Remdesivir doses, treatment cost, Neutrophil- Lymphocyte Ratio, Number of days hospitalized is store in data for further analysis.

Table 2 Total information gain (An Extremely Randomized tree classifier)

Feature	Value
Hospitalized	0.05
Number of days hospitalized	0.038
Medicines	0.03
Treatment costs	0.028
Neutrophil- Lymphocyte Ratio	0.026

In this study, we have extracted features through another method i.e. an extra-trees classifier. This class implements a meta-estimator that fits a number of randomized decision trees on various subsamples of our data and it uses averaging to improve the predictive accuracy and control over-fitting. The Gini function is used as supported criteria for identifying Gini impurity and entropy for the information gain is used to measure the quality of a split.

Table 2 explains the first five important features viz., hospitalized, number of days hospitalized, Medicines, treatment cost, Neutrophil Lymphocyte Ratio. Extra tree classifier features extracted five important features hospitalized, Number of days hospitalized, Medicines, treatment cost,

Neutrophil Lymphocyte Ratio important Features are selected by combining both methods.

Here, we have considered the combined features obtained from above both of the methods. The sample data contains six important features: hospitalized, number of days hospitalized, Neutrophil Lymphocyte Ratio, medicines, Remdesivir doses, treatment cost. Based on the above feature extraction further all the obtained specific features are used to find the accurate classification with Decision tree, Random forest, Logistic regression, neural network, SVM, KNN. Table 3-5 represents the results obtained from our proposed algorithm steps 8-10.

Table 3 Data Classification results (Chi-square feature extraction)

Classifier	Accuracy Score	Confusion Matrix	Metrics Classification Report			
			Score	Precision	Recall	F1-Score
Decision Tree	0.79	$\begin{bmatrix} 35 & 18 \\ 18 & 100 \end{bmatrix}$	0	0.66	0.66	0.66
			1	0.85	0.85	0.85
Random Forest	0.81	$\begin{bmatrix} 35 & 14 \\ 18 & 104 \end{bmatrix}$	0	0.71	0.66	0.69
			1	0.85	0.88	0.87
Logistic Regression	0.82	$\begin{bmatrix} 34 & 19 \\ 11 & 107 \end{bmatrix}$	0	0.76	0.64	0.69
			1	0.85	0.91	0.88
Neural Network MLP	0.92	$\begin{bmatrix} 43 & 10 \\ 3 & 115 \end{bmatrix}$	0	0.93	0.81	0.87
			1	0.92	0.97	0.95
Support Vector Machine	0.89	$\begin{bmatrix} 40 & 13 \\ 6 & 112 \end{bmatrix}$	0	0.87	0.75	0.81
			1	0.90	0.95	0.92
K Nearest Neighbor	0.78	$\begin{bmatrix} 29 & 24 \\ 13 & 105 \end{bmatrix}$	0	0.69	0.55	0.61
			1	0.81	0.89	0.85

Table 3 explains the features extracted through Chi-square method and results obtained from classifier techniques. It is observed that, neural network MLP gives 92 % accuracy, high precision and high F1-score as compared to other classifier techniques.

Table 4 Data Classification results (extra-trees classifier extraction)

Classifier	Accuracy Score	Confusion Matrix	Metrics Classification Report			
			Score	Precision	Recall	F1-Score
Decision Tree	0.79	$\begin{bmatrix} 29 & 11 \\ 24 & 107 \end{bmatrix}$	0	0.72	0.55	0.62
			1	0.82	0.91	0.86
Random Forest	0.78	$\begin{bmatrix} 27 & 11 \\ 26 & 107 \end{bmatrix}$	0	0.71	0.51	0.59
			1	0.80	0.91	0.85
Logistic Regression	0.80	$\begin{bmatrix} 31 & 22 \\ 11 & 107 \end{bmatrix}$	0	0.74	0.58	0.65
			1	0.83	0.91	0.87
Neural Network MLP	1.0	$\begin{bmatrix} 53 & 0 \\ 0 & 118 \end{bmatrix}$	0	1.00	1.00	1.00
			1	1.00	1.00	1.00
Support Vector Machine	0.89	$\begin{bmatrix} 40 & 13 \\ 6 & 112 \end{bmatrix}$	0	0.89	0.75	0.82
			1	0.90	0.96	0.93
K Nearest Neighbor	0.73	$\begin{bmatrix} 23 & 30 \\ 17 & 101 \end{bmatrix}$	0	0.57	0.43	0.49
			1	0.77	0.86	0.81

Table 4 explains the features extracted through Chi-square method and results obtained from classifier techniques. It is observed that, neural network MLP gives 100 % accuracy, high precision and high F1-score as compared to other classifier techniques.

Table 5 Features obtained from Chi-square + extra-trees classifier extraction

Classifier	Accuracy Score	Confusion Matrix	Metrics Classification Report			
			Score	Precision	Recall	F1-Score
Decision Tree	0.81	$\begin{bmatrix} 30 & 19 \\ 23 & 109 \end{bmatrix}$	0	0.77	0.57	0.65
			1	0.83	0.92	0.87
Random Forest	0.80	$\begin{bmatrix} 29 & 9 \\ 24 & 109 \end{bmatrix}$	0	0.76	0.55	0.64
			1	0.82	0.92	0.87
Logistic Regression	0.80	$\begin{bmatrix} 31 & 22 \\ 13 & 105 \end{bmatrix}$	0	0.70	0.58	0.64
			1	0.83	0.89	0.86
Neural Network MLP	1.0	$\begin{bmatrix} 53 & 0 \\ 0 & 118 \end{bmatrix}$	0	1.00	1.00	1.00
			1	1.00	1.00	1.00
Support Vector Machine	0.90	$\begin{bmatrix} 40 & 13 \\ 4 & 114 \end{bmatrix}$	0	0.91	0.75	0.82
			1	0.90	0.97	0.93
K Nearest Neighbor	0.80	$\begin{bmatrix} 30 & 23 \\ 12 & 106 \end{bmatrix}$	0	0.71	0.57	0.63
			1	0.82	0.90	0.86

Table 5 explains the features extracted through Chi-square method and results obtained from classifier techniques. It is observed that, neural network MLP gives 100 % accuracy, high precision and high F1-score as compared to other classifier techniques.

4. CONCLUSION

In this study, our main focus is to classify Covid-19 patients' recovery days based on various parameters. This study reveals that, 58.47 % of the respondents having fever and interesting fact is 12.58% of the overall respondents do not have any symptoms but they found to be Covid-19 positive. Also, 75 % of respondents do not require hospitalization. Initially, we have used 74 parameters for the study. Further, feature selection techniques were employed to identify the significant parameters. We have considered three different approaches: Chi-square method, extra tree classifier and third one is proposed as a combination of these methods features. Our analysis expressed that machine learning algorithms strengthen the analytical accuracy and the discriminative effectiveness of these classification. ML applications show potential results with high accuracy, sensitivity, and specificity using different models. The probability of accuracy has been analyzed from the characteristics obtained from the large feature set. The experimental results show the high accuracy for classifying the COVID-19 patient's recovery days with extracted features. The comparative results based on accuracy, precision, recall and F1-score measures. Classification performed with the MLP Neural network algorithm gives more suitable results as compared with other ML techniques. In a nutshell, the paper results explored supervised learning to detect the COVID-19 recovery patient's cases and neural network is an efficient algorithm as compared with other algorithms (100%). Hence, in future it is strongly recommended that for similar studies one can use ANN supervised learning for efficient estimation. This result gives key aspects to the administration and wellbeing establishment with the essential information that helps in resolution and development.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] A.M.U.D. Khanday, S.T. Rabani, Q.R. Khan, N. Rouf, M. Mohi Ud Din, Machine learning based approaches for detecting COVID-19 using clinical text data, Int. J. Inf. Tecnol. 12 (2020), 731–739.

[2] A. Pyrros, A. E. Flanders, J. M. Rodríguez-Fernández, et al. Predicting prolonged hospitalization and supplemental oxygenation in patients with COVID-19 infection from ambulatory chest radiographs using deep learning. Acad. Radiol. 28 (2021), 1151-1158.

[3] A. S. Kwekha-Rashid, H. N. Abduljabbar, B. Alhayani, Coronavirus disease (COVID-19) cases analysis using machine-learning applications. Appl. Nanosci. (2021). https://doi.org/10.1007/s13204-021-01868-7.

[4] T. Aishwarya, V. Ravi Kumar, Machine learning and deep learning approaches to analyze and detect COVID-19: A review, SN Comput. Sci. 2 (2021), 226.

[5] C.F.G. dos Santos, L.A. Passos, M.C. de Santana, J.P. Papa, Normalizing images is good to improve computer-assisted COVID-19 diagnosis, in: Data Science for COVID-19, Elsevier, 2021: pp. 51–62.

[6] C. Shorten, T.M. Khoshgoftaar, B. Furht, Deep learning applications for COVID-19, J. Big Data. 8 (2021), 18.

[7] D. Assaf, Y. Gutman, Y. Neuman, et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19, Int. Emerg. Med. 15 (2020), 1435–1443.

[8] E. Ayyildiz, M. Erdogan, A. Taskin, Forecasting COVID-19 recovered cases with Artificial Neural Networks to enable designing an effective blood supply chain, Computers Biol. Med. 139 (2021) 105029.

[9] E. Gambhir, R. Jain, A. Gupta, U. Tomer, Regression analysis of COVID-19 using machine learning algorithms, in: 2020 International Conference on Smart Electronics and Communication (ICOSEC), IEEE, Trichy, India, 2020: pp. 65–71.

[10] E. Karaarslan, D. Aydın, An artificial intelligence–based decision support and resource management system for COVID-19 pandemic, in: Data Science for COVID-19, Elsevier, 2021: pp. 25–49.

[11] F. Bekele, D.F. Mechessa, B. Sefera, Prevalence and associated factors of the psychological impact of COVID-19 among communities, health care workers and patients in Ethiopia: A systematic review, Ann. Med. Surgery. 66 (2021), 102403.

[12] F.N. Bernardo Gois, A. Lima, K. Santos, et al. Predictive models to the COVID-19, in: Data Science for COVID-19, Elsevier, 2021: pp. 1–24.

[13] A. Jafari-sales, H. Khaneshpour, M. Pashazadeh, R. Nasiri, Coronavirus Disease 2019 (COVID-19): review study, Jorjani Biomed J. 8 (2020), 4–10.

[14] H. Yasar, M. Ceylan, A novel comparative study for detection of Covid-19 on CT lung images using texture analysis, machine learning, and deep learning methods, Multimed. Tools Appl. 80 (2021), 5423–5447.

[15] https://www.calculator.net/sample-size-calculator.html

[16] J. Rasheed, A.A. Hameed, C. Djeddi, A. Jamil, F. Al-Turjman, A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images, Interdiscip. Sci. Comput. Life Sci. 13 (2021), 103–117.

[17] K. Munir, H. Elahi, M.U. Farooq, S. Ahmed, F. Frezza, A. Rizzi, Detection and screening of COVID-19 through chest computed tomography radiographs using deep neural networks., in: Data Science for COVID-19, Elsevier, 2021: pp. 63–73.

[18] P. Garrido, A.A. Adjei, J. Bajpai, et al. Has COVID-19 had a greater impact on female than male oncologists? Results of the ESMO Women for Oncology (W4O) Survey, ESMO Open. 6 (2021), 100131.

[19] R. Gupta, G. Pandey, P. Chaudhary, S.K. Pal, Machine Learning Models for Government to Predict COVID-19 Outbreak, Digit. Gov.: Res. Pract. 1 (2020), 1–6.

[20] R. O. Ogundokun, J.B. Awotunde. Machine learning prediction for covid 19 pandemic in India. medRxiv 2020.05.20.20107847; https://doi.org/10.1101/2020.05.20.20107847.

[21] R. Rajagopal, Comparative analysis of COVID-19 X-ray images classification using convolutional neural network, transfer learning, and machine learning classifiers using deep features, Pattern Recognit. Image Anal. 31 (2021), 313–322.

[22] R.S. Yadav, Data analysis of COVID-2019 epidemic using machine learning methods: a case study of India, Int. J. Inf. Tecnol. 12 (2020), 1321–1330.

[23] S. Malla, A. P.J.A., COVID-19 outbreak: An ensemble pre-trained deep learning model for detecting informative tweets, Appl. Soft Comput. 107 (2021), 107495.

[24] S. Ramanathan, M. Ramasundaram, Accurate computation: COVID-19 rRT-PCR positive test dataset using stages classification through textual big data mining with machine learning, J. Supercomput. 77 (2021), 7074–7088.

[25] T. Rogier, I. Eberl, F. Moretto, et al. COVID-19 or not COVID-19? Compared characteristics of patients hospitalized for suspected COVID-19, Eur. J. Clin. Microbiol. Infect. Dis. 40 (2021), 2023–2028.

[26] V. Chaurasia, S. Pal, Application of machine learning time series analysis for prediction COVID-19 pandemic, Res. Biomed. Eng. (2020). https://doi.org/10.1007/s42600-020-00105-4.

[27] Y. Quintero, D. Ardila, E. Camargo, F. Rivas, J. Aguilar, Machine learning models for the prediction of the SEIRD variables for the COVID-19 pandemic based on a deep dependence analysis of variables, Computers Biol. Med. 134 (2021), 104500.

[28] Z. Malki, E.-S. Atlam, A. Ewis, et al. The COVID-19 pandemic: prediction study based on machine learning models, Environ. Sci. Pollut. Res. 28 (2021), 40496–40506.