Original Article

Analysis and Comparison for Innovative Prediction Technique of COVID-19 using Logistic Regression algorithm over Support Vector Machine Algorithm with Improved Accuracy.

Garudadri Venkata Sree Charan[1] , Neelam Sanjeev Kumar [2*]

[1]Research Scholar,Department of Biomedical Engineering,Saveetha School Of Engineering,Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, 602105.
[2]Project Guide, Department of Biomedical Engineering,Saveetha School of Engineering,Saveetha Institute of Medical and Technical Sciences,Saveetha University, Chennai, Tamilnadu, India, 602105.

Abstract

Aim: The main objective of this study is to improve the accuracy of COVID-19 prediction and evaluation. Materials and Methods: This work depends on the data segregated from Kaggle's website where the samples are divided into two groups. Each group contains 20 samples (N=20) for both the Logistic regression and Support vector machine algorithms in accordance with the total sample size calculated using clinicalc.com by keeping alpha error-threshold at 0.05, confidence interval at 95%, enrolment ratio as 0:1, and G power at 80%. This involves training the data with validating 20 validations ranging from 5 to 24 in MatLab 2021a.
Results: The accuracy, sensitivity, and precision rates are compared using the SPSS Software and Independent sample T-test. The Logistic regression has better accuracy, sensitivity, and precision of 95.98%,94,65%, 96.20% (P<0.001) respectively compared to the Support vector machine where 91.25% of accuracy (P<0.001), 93.93% of sensitivity (P<0.001), and 86.11% of precision (P<0.001).
Conclusion: The Logistic regression algorithm produces superior outcomes than the Support vector machine algorithm.

Keywords: Innovative COVID-19 prediction, Machine learning, Logistic regression, Support vector machine, Accuracy.

DOI: 10.47750/pnr.2022.13.S03.050

INTRODUCTION

Coronaviruses are big RNA viruses that have been around since the 1960s. They cause upper respiratory tract infections that are mild to moderate, akin to the common cold (Paules, Marston, and Fauci 2020). The goal of this study was to create and compare prognosis prediction machine learning models based on invasive laboratory, non-invasive clinical, and demographic data collected from hospitals by using three Support vector machine models for comparison (Mahdavi et al. 2021). In this work, essential blood indices were retrieved from clinically accessible blood test data for proper identification of COVID-19 with comparable CT information and symptoms by a Random forest algorithm (Johri, Verma, and Paul 2020). The main application of this study is to use artificial intelligence in the COVID-19 chest X-Ray imaging diagnostic (Hassanien, Dey, and Elghamrawy 2020)(Sun et al. 2020).

Around 10 IEEE Explore and 80 ScienceDirect articles were observed to be identified with this work, which was finished in recent years and reported the developed algorithm and models using machine learning algorithms such as Naive Bayes, Decision tree, and neural network to predict and analyze innovative COVID-19 prediction performance in terms of accuracy, sensitivity, and precision. This article provides an overview of CNN increased performance with an average mean absolute percentage error of 3.718 %, among other things, in a comparative evaluation of machine learning techniques for COVID-19 transmission (Dairi et al. 2021). This study found that the ensemble bagged model has the greatest accuracy of 95.2%, 87.8% sensitivity, and 97% specificity among the three algorithms SVM, KNN, and EBM (Tamal et al. 2021). Based on several risk variables and demographic data, researchers estimated the criticality and survival chances of individuals with severe COVID-19 infection. The model's capacity to estimate mortality risk with 0.95 precision and 0.90 prediction accuracy were one of the main study's findings (Sardar, Sharma, and Gupta 2021). To forecast the severe instances of COVID-19 patients, researchers built a prediction model utilizing the support vector machine (SVM). The suggested model was proven to be effective and robust in predicting patients in severe circumstances with an accuracy of up to 0.775 (Soman, Loganathan, and Ajay 2009)

Our team has extensive knowledge and research experience that has translate into high quality publications (Bhansali et al. 2021; Jayanth et al. 2021; Sudhakar, Ravel, and Perumal 2021; Sathiyamoorthi et al. 2021; Deepanraj et al. 2021; Raju et al. 2021; Arun Prakash et al. 2020; Kamath et al. 2020; Shanmugam et al. 2021; Rajasekaran et al. 2020; Adhinarayanan et al. 2020; Rajesh et al. 2020; Aurtherson et al. 2021). The lack of a reliable early COVID-19 diagnostic that eliminates human error is the motivating factor for this study to predict COVID-19 at an early stage. The writers were machine learning specialists who were able to conduct biological research using the COVID-19 data and the Logistic regression and Support vector machine methods. The main objective is to identify COVID-19 with the greatest degree of accuracy possible.

MATERIALS AND METHODS

The study was carried out at the University simulation lab, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai. The sample size was determined using clinical.com, using an alpha error-threshold of 0.05, enrolment ratio of 0:1, 95 percent, based on earlier study findings (An et al. 2020). A Logistic regression algorithm (N=20) and SVM (N=20) were in Group 1. This research includes a total of 40 samples.

The Kaggle website provided the data samples utilized in the investigation. To acquire the absolute data necessary, the data set is subjected to data reduction procedures. To execute classification learning techniques, the data be fed into MatLab 2021a. To train, input data should be loaded into categorization learning systems. The imported data is trained twice, once for the Logistic regression with cross-validations ranging from 5 to 24 and once for the Support vector machine with cross-validations ranging from 5 to 24. After validation of data for an algorithm, the confusion matrix should be obtained for each validation (Dalrymple 2019), which involves the TP (true positive), TN (true negative), FP (false positive), FN (false negative). Accuracy, Sensitivity, and Precision are calculated with the help of these values given in Equation (1), (2), and (3).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

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

STATISTICAL ANALYSIS

IBM SPSS 27.0.1 was used to compare the accuracy of the Decision tree algorithm with the Support Vector Machine algorithm. The variables like COVID samples are independent and parameters like Asthma, Headache, Diabetes, Chronic Lung disease are dependent variables. The sample T-Test was performed to find the mean accuracy, mean sensitivity, and mean precision between the two groups, and performance comparison between the two groups is performed.

RESULTS

With an accuracy percentage ranging from 91.25 percent to 95.98 percent, sensitivity percentage ranging from 93.93 percent to 94.65 percent, and precision percentage ranging from 86.11 percent to 96.20 percent, both the methods appear to produce the same results as shown in Table 1a and Table 1b. When compared to the Support vector machine method, the Logistic regression approach exhibited the best accuracy, sensitivity, and precision as shown in Table 2. The Logistic regression has a lower error rate than the Support vector machine method, according to the statistical study in Table 2. Table 3 indicates that using the independent sample T-test, there appears to be a statistically insignificant difference (P=0.618 for accuracy, P=0.920 for sensitivity, P=0.40 for precision, $p<0.001$) in both techniques. These findings indicated that the Logistic regression outperforms the Support vector machine in predicting COVID-19 illness. Figure 1. represents the comparison of mean accuracy, sensitivity, and precision of innovative COVID-19 prediction with the Logistic regression algorithm and the Support vector machine algorithm. Figure 2a. and Figure 2b. represents the Confusion Matrix of Decision tree and Support Vector Machine which gives the true positive, true negative, false positive, and false negative values which are used for calculating the accuracy, sensitivity, and precision.

DISCUSSION

The Logistic regression method has the greatest accuracy (95.98 percent), sensitivity (94.65 percent), and precision (96.20 percent) shown in Table 2. Although not statistically significant, the significant difference appears to have risen slightly as shown in Table 3. The Logistic regression algorithm is the most straightforward and cost-effective method for innovative prediction of COVID-19.

The recent works are (Maggu and Singh 2021) In this study using statistical approaches such as univariate and multivariate regression methods, a study was done to determine the major variables that impact human infection with Logistic regression and SVM models to estimate the p-value (John and Shaiba 2019). The research was carried out to enhance medical diagnosis algorithms for binary and multiclass issues in MERS-CoV datasets, the Decision tree classifier achieves a maximum accuracy of 86.44% compared to the SVM method (AlMoammar, AlHenaki, and Kurdi 2019). The author used three machine learning approaches to predict MERS CoV by Naive Bayes, Random forest, SVM whose accuracies are 94.2%, 90%, and 68% respectively (Li and Du 2019). This research looks at how demographics, vital signs, and chronic illness may be used to predict COVID-19-positive individuals' survival and death. With an accuracy of 0.95 and an AUC value of 0.99, the Random forest outperformed the other methods (Aljameel et al. 2021). In this study, the authors analyzed COVID-19 data sets to determine the most affected age group with help of different machine learning algorithms like Decision tree classifier, SVM, KNN, XGBoost Classifier, etc. (Jain and Chatterjee 2020).

Age, gender, and whether or not the patient has a condition or sickness such as diabetes, pneumonia, asthma, obesity, or cardiac issues are all factors that impact the research. Because the data will not be in time series, this will be a big problem. This article looks at how COVID-19 detection technology is utilized in the healthcare business and how it can aid with more accurate diagnosis in the near future. As a result, this effort has a bright future since manual forecasting may easily be converted to automated output at a low cost. A larger dataset of real-time applications, paired with more machine learning algorithms, might produce better outcomes. The limitations of this research work are to increase more sample size by capitulating significant accuracy than the existing algorithm in the Innovative detection model and the future scope of this research is to ensemble the simple genetic algorithm in predicting all variants of COVID and classifying the Adaboost for feature extraction.

CONCLUSION

When compared to SVM, the Matlab-based Logistic regression algorithm (95.98 percent) produced better results in this innovative COVID-19 prediction study for the Support vector machine (91.25 percent). Furthermore, unlike previous approaches, the algorithm's performance improved as the amount of data increased. This model is quite efficient and shows a lot of potential when it comes to predicting and assessing COVID-19, thus it may be utilized in hospitals and testing facilities.

DECLARATION

Conflicts of Interest

No conflict of interest in this manuscript.

Author Contributions

Author GVSC was involved in data collection, data analysis & manuscript writing. The author's learning-Based guide NSK was involved in conceptualization, data validation, and critical review of manuscripts.

Acknowledgment

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for successfully carrying out this work.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1.Venus Electronics, Chennai
2.Saveetha University
3.Saveetha Institute of Medical And Technical Sciences
4.Saveetha School of Engineering

REFERENCES

1. Adhinarayanan, Rajesh, Aravindh Ramakrishnan, Gopal Kaliyaperumal, Melvinvíctor De Poures, Rajesh Kumar Babu, and Damodharan Dillikannan. 2020. “Comparative Analysis on the Effect of 1-Decanol and Di-N-Butyl Ether as Additive with diesel/LDPE Blends in Compression Ignition Engine.” *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, June, 1–18.
2. Aljameel, Sumayh S., Irfan Ullah Khan, Nida Aslam, Malak Aljabri, and Eman S. Alsulmi. 2021. “Machine Learning-Based Model to Predict the Disease Severity and Outcome in COVID-19 Patients.” *Scientific Programming*. https://doi.org/10.1155/2021/5587188.
3. AlMoammar, Afnan, Lubna AlHenaki, and Heba Kurdi. 2019. “Selecting Accurate Classifier Models for a MERS-CoV Dataset.” *Advances in Intelligent Systems and Computing*. https://doi.org/10.1007/978-3-030-01054-6_74.
4. An, Chansik, Hyunsun Lim, Dong-Wook Kim, Jung Hyun Chang, Yoon Jung Choi, and Seong Woo Kim. 2020. “Machine Learning Prediction for Mortality of Patients Diagnosed with COVID-19: A Nationwide Korean Cohort Study.” *Scientific Reports* 10 (1): 18716.
5. Arun Prakash, V. R., J. Francis Xavier, G. Ramesh, T. Maridurai, K. Siva Kumar, and R. Blessing Sam Raj. 2020. “Mechanical, Thermal and Fatigue Behaviour of Surface-Treated Novel Caryota Urens Fibre–reinforced Epoxy Composite.” *Biomass Conversion and Biorefinery*, August. https://doi.org/10.1007/s13399-020-00938-0.
6. Aurtherson, P. Babu, Bhanu Teja Nalla, Karthikeyan Srinivasan, Kulmani Mehar, and Yuvarajan Devarajan. 2021. “Biofuel Production from Novel Prunus Domestica Kernel Oil: Process Optimization Technique.” *Biomass Conversion and Biorefinery*, May. https://doi.org/10.1007/s13399-021-01551-5.
7. Bhansali, Karan J., Kamlesh R. Balinge, Subodh U. Raut, Shubham A. Deshmukh, M. Senthil Kumar, C. Ramesh Kumar, and Pundlik R. Bhagat. 2021. “Visible Light Assisted Sulfonic Acid-Functionalized Porphyrin Comprising Benzimidazolium Moiety for Photocatalytic Transesterification of Castor Oil.” *Fuel* 304 (November): 121490.
8. Dairi, Abdelkader, Fouzi Harrou, Abdelhafid Zeroual, Mohamad Mazen Hittawe, and Ying Sun. 2021. “Comparative Study of Machine Learning Methods for COVID-19 Transmission Forecasting.” *Journal of Biomedical Informatics* 118 (June): 103791.
9. Dalrymple, Theodore. 2019. *False Positive: A Year of Error, Omission, and Political Correctness in the New England Journal of Medicine*. Encounter Books.
10. Deepanraj, B., N. Senthilkumar, D. Mala, and A. Sathiamourthy. 2021. “Cashew Nut Shell Liquid as Alternate Fuel for CI Engine—optimization Approach for Performance Improvement.” *Biomass Conversion and Biorefinery*, February. https://doi.org/10.1007/s13399-021-01312-4.
11. Hassanien, Aboul-Ella, Nilanjan Dey, and Sally Elghamrawy. 2020. *Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach*. Springer Nature.
12. Jain, Vishal, and Jyotir Moy Chatterjee. 2020. *Machine Learning with Health Care Perspective: Machine Learning and Healthcare*. Springer Nature.
13. Jayanth, Bellappu Venkat, Melvin Victor Depoures, Gopal Kaliyaperumal, Damodharan Dillikannan, Dilipsingh Jawahar, Kumaran Palani, and Ganesha Prasad Meravanigee Shivappa. 2021. “A Comprehensive Study on the Effects of Multiple Injection Strategies and Exhaust Gas Recirculation on Diesel Engine Characteristics That Utilize Waste High Density Polyethylene Oil.” *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, June, 1–18.
14. John, Maya, and Hadil Shaiba. 2019. “Main Factors Influencing Recovery in MERS Co-V Patients Using Machine Learning.” *Journal of Infection and Public Health* 12 (5): 700–704.
15. Johri, Prashant, Jitendra Kumar Verma, and Sudip Paul. 2020. *Applications of Machine Learning*. Springer Nature.
16. Kamath, Manjunath, Subha Krishna Rao, Jaison, Sridhar, Kasthuri, Gopinath, Sivaperumal, and Shantanu Patil. 2020. “Melatonin Delivery from PCL Scaffold Enhances Glycosaminoglycans Deposition in Human Chondrocytes – Bioactive Scaffold Model for Cartilage Regeneration.” *Process Biochemistry* 99 (December): 36–47.
17. Li, Fang, and Lanying Du. 2019. *MERS-CoV*. MDPI.
18. Maggu, Sarthak, and Vijander Singh. 2021. “Prediction and Analysis of Covid-19 with Various Machine Learning Algorithms.” *Journal of Intelligent Systems and Computing*. https://doi.org/10.51682/jiscom.00202004.2021.
19. Mahdavi, Mahdi, Hadi Choubdar, Erfan Zabeh, Michael Rieder, Safieddin Safavi-Naeini, Zsolt Jobbagy, Amirata Ghorbani, et al. 2021. “A Machine Learning Based Exploration of COVID-19 Mortality Risk.” *PloS One* 16 (7): e0252384.
20. Paules, Catharine I., Hilary D. Marston, and Anthony S. Fauci. 2020. “Coronavirus Infections—More Than Just the Common Cold.” *JAMA*. https://doi.org/10.1001/jama.2020.0757.
21. Rajasekaran, S., D. Damodharan, K. Gopal, B. Rajesh Kumar, and Melvin Victor De Poures. 2020. “Collective Influence of 1-Decanol Addition, Injection Pressure and EGR on Diesel Engine Characteristics Fueled with diesel/LDPE Oil Blends.” *Fuel* 277 (October): 118166.
22. Rajesh, A., K. Gopal, De Poures Melvin Victor, B. Rajesh Kumar, A. P. Sathiyagnanam, and D. Damodharan. 2020. “Effect of Anisole Addition to Waste Cooking Oil Methyl Ester on Combustion, Emission and Performance Characteristics of a DI Diesel Engine without Any Modifications.” *Fuel* 278 (October): 118315.
23. Raju, P., K. Raja, K. Lingadurai, T. Maridurai, and S. C. Prasanna. 2021. “Glass/Caryota Urens Hybridized Fibre-Reinforced nanoclay/SiC Toughened Epoxy Hybrid Composite: Mechanical, Drop Load Impact, Hydrophobicity and Fatigue Behaviour.” *Biomass Conversion and Biorefinery*, March. https://doi.org/10.1007/s13399-021-01427-8.
24. Sardar, Rahila, Arun Sharma, and Dinesh Gupta. 2021. “Machine Learning Assisted Prediction of Prognostic Biomarkers Associated With COVID-19, Using Clinical and Proteomics Data.” *Frontiers in Genetics*. https://doi.org/10.3389/fgene.2021.636441.
25. Sathiyamoorthi, Ramalingam, Gomathinayakam Sankaranarayanan, Dinesh Babu Munuswamy, and Yuvarajan Devarajan. 2021.

"Experimental Study of Spray Analysis for Palmarosa Biodiesel-diesel Blends in a Constant Volume Chamber." *Environmental Progress & Sustainable Energy* 40 (6). https://doi.org/10.1002/ep.13696.

26. Shanmugam, Rajasekaran, Damodharan Dillikannan, Gopal Kaliyaperumal, Melvin Victor De Poures, and Rajesh Kumar Babu. 2021. "A Comprehensive Study on the Effects of 1-Decanol, Compression Ratio and Exhaust Gas Recirculation on Diesel Engine Characteristics Powered with Low Density Polyethylene Oil." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* 43 (23): 3064–81.
27. Soman, K. P., R. Loganathan, and V. Ajay. 2009. *Machine Learning with SVM and Other Kernel Methods*. PHI Learning Pvt. Ltd.
28. Sudhakar, M. P., Merlyn Ravel, and K. Perumal. 2021. "Pretreatment and Process Optimization of Bioethanol Production from Spent Biomass of Ganoderma Lucidum Using Saccharomyces Cerevisiae." *Fuel* 306 (December): 121680.
29. Sun, Liping, Fengxiang Song, Nannan Shi, Fengjun Liu, Shenyang Li, Ping Li, Weihan Zhang, et al. 2020. "Combination of Four Clinical Indicators Predicts the Severe/critical Symptom of Patients Infected COVID-19." *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology* 128 (July): 104431.
30. Tamal, Mahbubunnabi, Maha Alshammari, Meernah Alabdullah, Rana Hourani, Hossain Abu Alola, and Tarek M. Hegazi. 2021. "An Integrated Framework with Machine Learning and Radiomics for Accurate and Rapid Early Diagnosis of COVID-19 from Chest X-Ray." *Expert Systems with Applications* 180 (October): 115152.

TABLES AND FIGURES

Table 1a. Covid-19 samples using Logistic Regression

Sample	Accuracy	Sensitivity	Precision
1	0.95	0.94	0.94
2	0.95	0.94	0.94
3	0.97	0.94	1
4	0.95	0.94	0.94
5	0.95	0.94	0.94
6	0.95	0.94	0.94
7	0.95	0.94	0.94
8	0.95	0.94	0.94
9	0.97	0.94	0.94
10	0.98	0.95	1
11	0.95	0.94	0.94
12	0.95	0.94	0.94
13	0.95	0.94	0.94
14	0.97	0.94	1
15	0.95	0.94	0.94
16	0.94	0.94	0.94
17	0.97	0.94	1
18	0.95	0.94	0.94
19	0.97	0.94	1
20	0.97	0.94	1

Table 1b. Covid-19 samples using Support Vector Machine

Sample	Accuracy	Sensitivity	Precision
1	0.9	0.93	0.83
2	0.92	0.94	0.88
3	0.9	0.94	0.83
4	0.92	0.94	0.88
5	0.9	0.91	0.83
6	0.92	0.93	0.88
7	0.9	0.92	0.83
8	0.92	0.94	0.88
9	0.9	0.94	0.83
10	0.92	0.93	0.88
11	0.9	0.92	0.83
12	0.92	0.94	0.88
13	0.9	0.93	0.83
14	0.92	0.92	0.88
15	0.9	0.93	0.83
16	0.92	0.93	0.88
17	0.9	0.94	0.83
18	0.92	0.94	0.88
19	0.9	0.94	0.83
20	0.92	0.9	0.88

Table 2. Comparison of mean accuracy, sensitivity, and precision using Logistic regression and SVM algorithms

GROUP STATISTICS					
Parameters	**Group**	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Accuracy	Logistic Regression	20	0.9598	0.01321	0.00295
	Support Vector Machine	20	0.9125	0.01282	0.00287

Sensitivity	Logistic Regression	20	0.9465	0.00269	0.00060
	Support Vector Machine	20	0.9393	0.00189	0.00042
Precision	Logistic Regression	20	0.9620	0.02563	0.00573
	Support Vector Machine	20	0.8611	0.02850	0.00637

Table 3. Independent sample T-test in predicting the accuracy, sensitivity, and precision of COVID-19 using the Logistic regression and SVM algorithm. There appears to be an insignificant difference in both methods.

Parameter	**Equal Variances**	**Levene's Test for Equality of Variances**		**T-test for Equality of Means**					
		F	**Sig**	**t**	**df**	**Significance (one-Sided p)**	**Mean Difference**	**Std.Error Difference**	**95% Confidence interval (Upper)**
Accuracy	**Assumed**	0.252	0.618	11.499	38	<.001	0.4734	.00412	.03901
	Not assumed			11.499	37.96	<.001	0.4734	.00412	.03901
Sensitivity	Assumed	0.010	0.920	9.729	38	<.001	0.0071	.00074	.00566
	Not assumed			9.729	34.02	<.001	0.0071	.00074	.00566
Precision	Assumed	4.532	0.040	11.767	38	<.001	0.1008	.00857	.08349
	Not assumed			11.767	37.57	<.001	0.1008	.00857	.08349

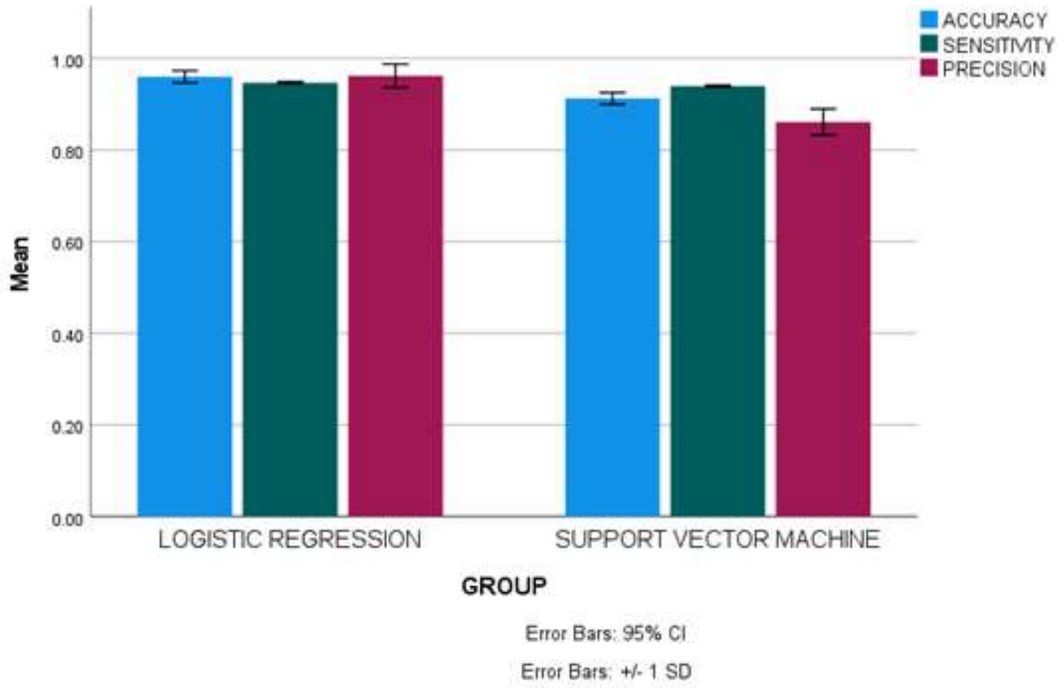

Fig. 1. Bar graph representing the comparison of mean accuracy, sensitivity, and precision of COVID-19 prediction with Logistic regression algorithm and the Support vector machine algorithm. Both the techniques appear to produce the same variable results with accuracy ranging from 95.98% to 91.25%. X-axis: Logistic regression vs SVM. Y-axis: mean accuracy, sensitivity, and precision detection ± 1 SD.

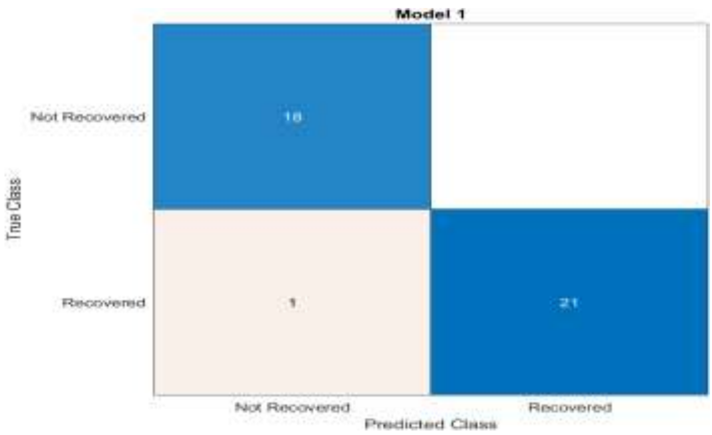

Fig. 2a. Confusion matrix of Logistic regression for K=5. True Positive is found to be 17% and false positive is found to be 1%, true negative is found to be 21% and false negative is found to be 1%.

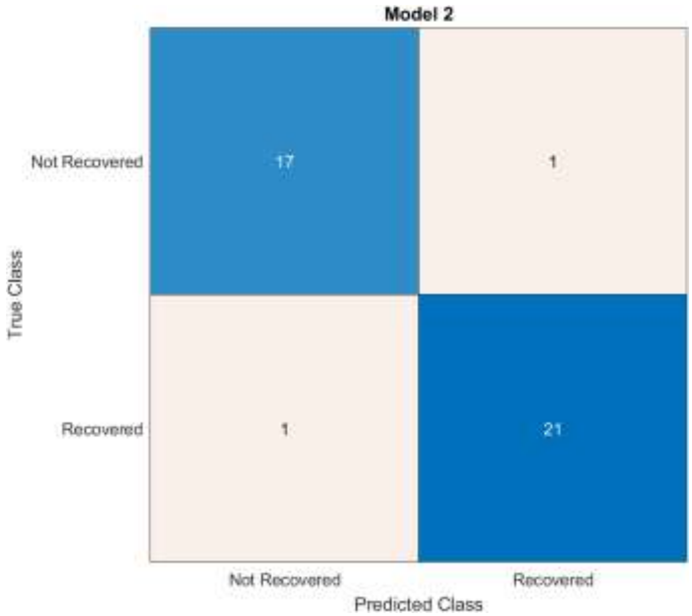

Fig. 2b. Confusion matrix for Support Vector Machine algorithm for K= 5. True Positive is found to be 17% and false positive is found to be 1%, true negative is found to be 21% and false negative is found to be 1%.