

Analysis and Comparison for Innovative Prediction Technique of COVID-19 using Decision Tree Algorithm over the Support Vector Machine Algorithm with Improved Accuracy

Garudadri Venkata Sree Charan¹, Neelam Sanjeev Kumar²

Research Scholar, Department of Biomedical Engineering, Saveetha School OF Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode: 602105.

ABSTRACT

Aim: The primary goal of this research is to increase the accuracy of COVID-19 prediction and its analysis. **Materials and Method:** This study relied on data collected from Kaggle's website and samples are divided into two groups, GROUP 1 (N=20) for the Decision tree and GROUP 2 (N=20) for the Support Vector Machine (SVM) in accordance with the total sample size calculated using clinical.com by keeping alpha error-threshold value 0.05, 95% confidence interval, enrolment ratio as 0:1, and G power at 80%. It involves the software implementation program in MatLab 2021a validating with 20 validations. **Results:** The accuracy, sensitivity, and precision rates are compared using Statistical Package for the Social Sciences (SPSS) software and an Independent sample T-Test. In comparison to the two, the Decision tree 93.91% accuracy, 94.33% sensitivity, 92% precision with $P=0.001$ ($p<0.05$) produces a superior outcome to the Support Vector Machine 91.25% accuracy, 93.93% sensitivity, 86.11% precision ($P<0.001$). **Conclusion:** The decision tree algorithm produces better results compared to the Support Vector Machine.

Keywords

Innovative COVID-19 prediction, Machine learning, Decision tree, Support vector machine, Accuracy.

Imprint

G Garudadri Venkata Sree Charan, Neelam Sanjeev Kumar. Analysis and Comparison for Innovative Prediction Technique of COVID-19 using Decision Tree Algorithm over the Support Vector Machine Algorithm with Improved Accuracy. *Cardiometry*; Special issue No. 25; December 2022; p. 891-896; DOI: 10.18137/cardiometry.2022.25.891896; Available

from: <http://www.cardiometry.net/issues/no25-december-2022/prediction-technique-covid-19>

INTRODUCTION

COVID-19 stands for a coronavirus that has yet to be thoroughly identified in humans. This condition is known to affect people of all ages, especially those with bodily diseases, high blood pressure, diabetes, heart difficulties, or compromised immune systems (Zagrouba et al. 2021). This paper presents a predictive framework combining Support Vector Machine in forecasting COVID-19 (Albahri et al. 2020). This study looked at state-of-the-art COVID prediction systems based on data mining and machine learning. A supervised machine learning model was constructed with the purpose of predicting the presence of COVID-19 in a person (Villavicencio et al. 2021). By efficiently identifying at-risk COVID-19 patients early, the suggested methodology can aid decision-makers and healthcare professionals (Aljameel et al. 2021) (Hunt, Cock, and Symonds 2021).

About 7 IEEE Explore and 74 ScienceDirect articles were found to be related to this work, which was completed in recent years and reported the developed algorithm and models using ML algorithms such as decision tree, SVM, logistic regression, and neural network to predict and analyze performance with regard to accuracy, sensitivity and precision in innovative COVID-19 projection (Kwekha-Rashid, Abduljabbar, and Alhayani 2021) where the purpose of this study was to figure out how machine learning applications and algorithms fit into the COVID-19 examination, among other things, with 92.9% testing accuracy Supervised learning outperformed rival Unsupervised learning methods. The main goal of this study was to develop and test machine learning models that could estimate the outcome of COVID-19 patients using demographic data, virus routines, and health conditions, the LASSO, and linear SVM gave better accuracy rates than the other algorithms (Gambhir et al. 2020). With COVID-19 and Pneumonia cases on lungs X-ray images, this study offers a machine learning-based classification of the extracted deep feature using ResNet152 (Diallo, Diallo, and Ju 2020) (Alballa and Al-Turaiki 2021) the model has an RF accuracy of 0.973 and an XGBoost predictive classifier accuracy of 0.977. This article presents an eight-question ML model for pre-

dicting a true SARS-Cov-2 virus through RT-PCR tests (Zoabi, Deri-Rozov, and Shomron 2021). Our team has extensive knowledge and research experience that has translate into high quality publications (Chellapa et al. 2020; Lavanya et al. 2021; Raj Ret al. 2020; Shilpa-Jain et al. 2021; S et al. 2021; Ramadoss et al. 2022; Wuet al. 2020; Kalidoss et al. 2021; Kaja et al. 2020; Antink et al. 2020; Paulet al. 2020; Malaikolundhan et al. 2020)

The lack of effective early diagnosis of COVID-19 that removes the human error rates is the driving force for this research to anticipate COVID-19 in a preliminary phase. The authors were experts in machine learning algorithms and they were able to perform biological research comparing the COVID-19 data with the Decision tree algorithm and Support Vector Machine. The major goal is to detect COVID-19 with the maximum degree of accuracy.

MATERIALS AND METHODS

The research was carried out at the University simulation lab, Saveetha School of Engineering, Chennai. The sample size was estimated using clinical.com and earlier study data (An et al. 2020) with an alpha error-threshold of 0.05, enrolment ratio of 0:1, 95 percent confidence interval, and power of 80 percent. Group 1 consisted of a Decision tree algorithm (N=20) and SVM (N=20). This research comprised a total of 40 samples.

Data samples used in this study are collected from the Kaggle website. The data set undergoes data reduction techniques to obtain the absolute data required. The data should be given as input to MatLab 2021a to perform classification learning techniques. Input data should be imported to classification learning tools to perform training. The imported data was trained separately for each algorithm i.e, once for the Decision tree varying with cross-validations from 5 to 24 and similarly for the Support Vector Machine with cross-validations from 5 to 24. After validation of data for an algorithm, the confusion matrix should be obtained for each validation (Hunt, Cock, and Symonds 2021), 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

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

RESULTS

Both the algorithms appear to provide the same COVID-19, with an accuracy rate ranging from 91.25 percent to 93.93 percent. The mean accuracy of the Decision Tree (DT) method is greater than that of the SVM approach, as shown in Table 2. The DT algorithm achieved the best accuracy, sensitivity, and precision when compared to the Support Vector Machine algorithm as shown in Table 1a and Table 1b. According to the statistical analysis of Table 2, the Decision tree has a lower error rate than the SVM. Table 3 shows that using an independent sample T-test, there appears to be a statistically negligible difference (P=0.038 for accuracy, P=0.20 with (p>0.05) for sensitivity and P=0.32 for precision, p>0.05, in both approaches. These findings revealed that the Decision tree algorithm can predict COVID-19 disease more quickly than the Support Vector Machine. The mean accuracy, sensitivity, and precision of the new COVID-19 prediction with the Decision tree method and the Support vector machine technique are shown in Fig. 1. The confusion matrix of DT and SVM algorithms, shown in Fig. 2a. and Fig. 2b., offers TP, TN, FP, FN values, which can be used to calculate accuracy, sensitivity, and precision.

DISCUSSION

In this study of innovative COVID-19 prediction, the Decision tree algorithm showed the best accuracy (93.31%), sensitivity (94.33%), precision (92%) as shown in Table 2. The significant difference appears to have increased slightly, although it is not statistically significant as shown in Table 3. Decision tree algorithm is the simplest and most cost-effective approach to predict COVID-19.

Table 1a

Covid-19 samples using Decision tree

Sample	Accuracy	Sensitivity	Precision
1	0.95	0.94	0.94
2	0.97	0.94	1
3	0.95	0.94	0.94
4	0.95	0.94	0.94
5	0.92	0.91	0.88
6	0.95	0.93	0.94
7	0.9	0.92	0.83
8	0.92	0.94	0.88
9	0.95	0.94	0.94
10	0.92	0.93	0.88
11	0.9	0.92	0.83
12	0.95	0.94	0.94
13	0.92	0.93	0.88
14	0.97	0.92	0.83
15	0.92	0.93	0.94
16	0.95	0.92	0.94
17	0.92	0.93	0.88
18	0.95	0.94	1
19	0.92	0.94	0.88
20	0.95	0.9	0.94

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 Decision tree and SVM algorithms.

GROUP STATISTICS					
Parameters	Group	N	Mean	Std. Deviation	Std. ErrorMean
Accuracy	DecisionTree	20	0.9391	0.02072	0.00463
	Support VectorMachine	20	0.9125	0.01282	0.00287
Sensitivity	DecisionTree	20	0.9433	0.00316	0.00071
	Support VectorMachine	20	0.9393	0.00189	0.00042
Precision	DecisionTree	20	0.92	0.04622	0.1033
	Support VectorMachine	20	0.8611	0.02850	0.00637

The researchers worked on SVM and Decision tree algorithms for innovative prediction of COVID-19 which results in 91% sensitivity of the SVM model (Alballa and Al-Turaiki 2021). In this study, the authors used a CR meta classifier and Decision tree for analyzing the early diagnosis of COVID which resulted in 82% to 86% of accuracy for the Decision tree (Arpaci et al. 2021). This study shows that among the three algorithms SVM model gave the highest accuracy of 95.2 percent, 87.8% sensitivity, 97% specificity (Tamal et al. 2021). Related work on SVM, Gradient Boosted Decision Tree are among the machine learning approaches used to create this ensemble model which produces an accuracy of 96.21% (Gao et al. 2020) (Li et al. 2021). Clinical prediction models are evaluated with the use of ML and laboratory data, and accuracy, F1-score, precision, and recall are 86.66 %, 91.89 %, 86.75 %, and 99.42 percent, respectively (Alakus and Turkoglu 2020).

The factors which influence the study are the attributes like age, gender, whether the patient is having an illness or disease like diabetes, pneumonia, asthma, obesity, heart problems, etc. As the data will not be in time series it will be the major issue faced. The future work is based on how COVID-19 detection technology is being used in the healthcare industry and how it may help with a more accurate diagnosis. As a consequence, this initiative has a promising future, as manual forecasting may be readily converted to computerized output at a cheap cost. A larger dataset of real-time applications combined with additional machine learning algorithms might yield superior results. The limitations of this research work are to increase

Table 3

Independent sample T-test in predicting the accuracy, sensitivity, and precision of COVID-19 using the Decision tree and SVM algorithm. There appears to be an insignificant difference in both methods for Accuracy and precision with $p > 0.05$

Parameter	Equal Vari- ances	Levene's Test for Equality of Variances		T-test for Equality of Means					
		F	Sig	t	df	Significance (one-Sided p)	Mean Dif- ference	Std.Error Difference	95% Confidence interval (Upper)
Accuracy	Assumed	4.647	0.38	4.87	38	<.001	.02659	.00545	.03762
	Not assumed			4.87	31.69	<.001	.02659	.00545	.03769
Sensitivity	Assumed	5.879	0.02	4.86	38	<.001	.00400	.00082	.00566
	Not assumed			4.86	31.02	<.001	.00400	.00082	.00568
Precision	Assumed	4.930	0.03	4.85	38	<.001	.05892	.01214	.08350
	Not assumed			4.85	31.62	<.001	.05892	.01214	.08366

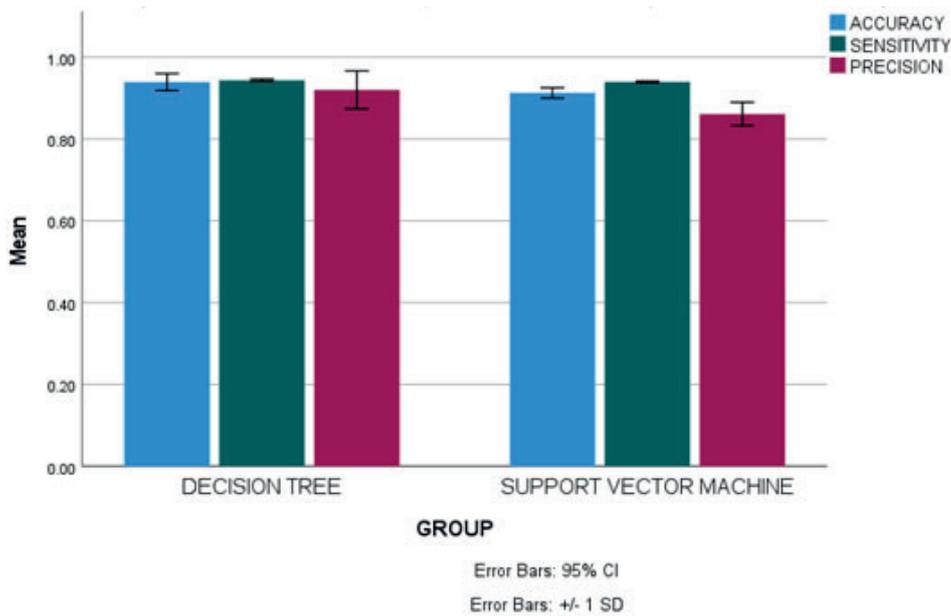


Fig. 1. Bar graph representing the comparison of mean accuracy, sensitivity, and precision of COVID-19 prediction with Decision tree algorithm and the Support vector machine algorithm. Both the techniques appear to produce the same variable results with accuracy ranging from 93% to 91%. X-axis: Decision tree vs SVM. Y-axis: mean accuracy, sensitivity, and precision detection \pm 1 SD.

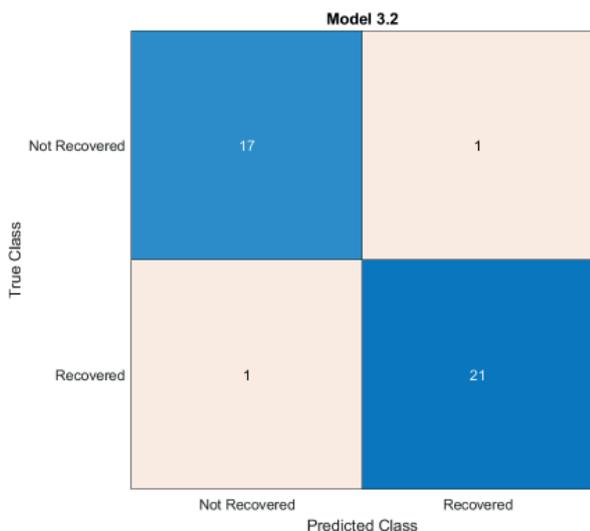


Fig. 2a. Confusion matrix for Decision tree algorithm K=5. True Positive is found to be 17% and false positive is found to be 1%, true negative is found to be 21% and false negative is found to be 1%.

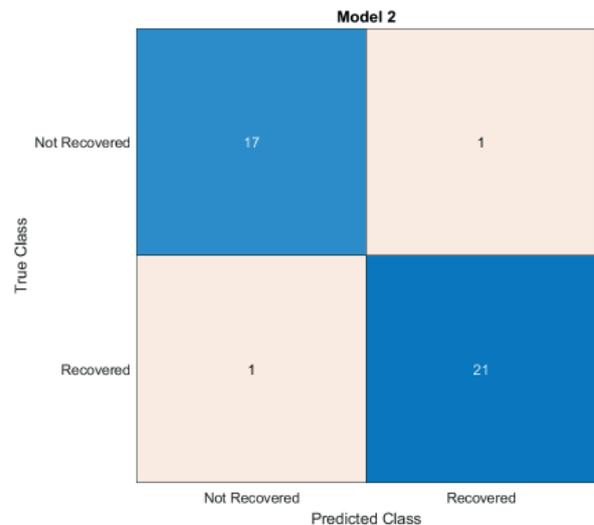


Fig. 2b. Confusion matrix for Support Vector Machine algorithm for K= 5. True Positive is found to be 17% and false positive is found to be 1%, true negative is found to be 21% and false negative is found to be 1%.

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 Decision tree algorithm (93.91%) that runs in the MatLab proved to offer higher results in this research of innovative COVID-19 prediction for Support vector machine (91.25%). In addition, the performance of the algorithm improved as the amount of the data grew, which is not seen in other methods. This model is highly efficient and has a lot of promise in terms of predicting and analyzing COVID-19, thus it may be used 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. Alakus, Talha Burak, and Ibrahim Turkoglu. 2020. "Comparison of Deep Learning Approaches to Predict

COVID-19 Infection." *Chaos, Solitons, and Fractals* 140 (November): 110120.

2. Albahri, A. S., Rula A. Hamid, Jwan K. Alwan, Z. T. Al-Qays, A. A. Zaidan, B. B. Zaidan, A. O. S. Albahri, et al. 2020. "Role of Biological Data Mining and Machine Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19): A Systematic Review." *Journal of Medical Systems* 44 (7): 122.

3. Alballa, Norah, and Isra Al-Turaiki. 2021. "Machine Learning Approaches in COVID-19 Diagnosis, Mortality, and Severity Risk Prediction: A Review." *Informatics in Medicine Unlocked* 24 (April): 100564.

4. 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* 2021 (April). <https://doi.org/10.1155/2021/5587188>.

5. 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.

6. Arpaci, Ibrahim, Shigao Huang, Mostafa Al-Emran, Mohammed N. Al-Kabi, and Minfei Peng. 2021. "Predicting the COVID-19 Infection with Fourteen Clinical Features Using Machine Learning Classification Algorithms." *Multimedia Tools and Applications*, January, 1–15.

7. Diallo, Papa Abdou Karim Karou, Papa Abdou Karim Diallo, and Yun Ju. 2020. "Accurate Detection of COVID-19 Using K-EfficientNet Deep Learning Image Classifier and K-COVID Chest X-Ray Images Dataset." 2020 IEEE 6th International Conference on Computer and Communications (ICCC). <https://doi.org/10.1109/iccc51575.2020.9344949>.

8. Gambhir, Ekta, Ritika Jain, Alankrit Gupta, and Uma Tomer. 2020. "Regression Analysis of COVID-19 Using Machine Learning Algorithms." 2020 International Conference on Smart Electronics and Communication (ICOSEC). <https://doi.org/10.1109/icosec49089.2020.9215356>.

9. Gao, Yue, Guang-Yao Cai, Wei Fang, Hua-Yi Li, Si-Yuan Wang, Lingxi Chen, Yang Yu, et al. 2020. "Machine Learning Based Early Warning System Enables Accurate Mortality Risk Prediction for COVID-19." *Nature Communications* 11 (1): 5033.

10. Hunt, J. S., C. Cock, and E. L. Symonds. 2021. "A True Positive and a False Negative? The Dilemma of

- Negative Colonoscopy After a Positive Fecal Occult Blood Test.” *Digestive Diseases and Sciences*. <https://doi.org/10.1007/s10620-021-06986-4>.
11. Kwekha-Rashid, Ameer Sardar, Heamn N. Abduljabbar, and Bilal Alhayani. 2021. “Coronavirus Disease (COVID-19) Cases Analysis Using Machine-Learning Applications.” *Applied Nanoscience*. <https://doi.org/10.1007/s13204-021-01868-7>.
 12. Li, Xue, Chiaki Ono, Noriko Warita, Tomoka Shoji, Takashi Nakagawa, Hitomi Usukura, Zhiqian Yu, et al. 2021. “Heart Rate Information-Based Machine Learning Prediction of Emotions Among Pregnant Women.” *Frontiers in Psychiatry / Frontiers Research Foundation* 12: 799029.
 13. 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.
 14. Villavicencio, Charlyn Nayve, Julio Jerison Escudero Macrohon, Xavier Alphonse Inbaraj, Jyh-Horng Jeng, and Jer-Guang Hsieh. 2021. “COVID-19 Prediction Applying Supervised Machine Learning Algorithms with Comparative Analysis Using WEKA.” *Algorithms*. <https://doi.org/10.3390/a14070201>.
 15. Zagrouba, Rachid, Muhammad Adnan Khan, Atta-ur-Rahman, Muhammad Aamer Saleem, Muhammad Faheem Mushtaq, Abdur Rehman, and Muhammad Farhan Khan. 2021. “Modelling and Simulation of COVID-19 Outbreak Prediction Using Supervised Machine Learning.” *Computers, Materials & Continua*. <https://doi.org/10.32604/cmc.2021.014042>.
 16. Zoabi, Yazeed, Shira Deri-Rozov, and Noam Shomron. 2021. “Machine Learning-Based Prediction of COVID-19 Diagnosis Based on Symptoms.” *NPJ Digital Medicine* 4 (1): 3.