COVID-19 Detection Using Symptoms

Authors : Chaitanya Yadavally, Chaitanya Veeragandhapu, Rohit Tripuraneni

Abstract

COVID-19 is a contagious disease which has mutated into many variants and that has become a global pandemic which demands diagnosis to be done soon. In this project, COVID-19 prediction is done using machine learning algorithms, here many algorithms will be compared in order to make a comparative analysis. Predicting COVID-19 will be done with COVID-19 symptoms using ML. Dataset is taken from a public repository. Many factors like cough, fever, headache, contact, shortness of breath etc.. are taken as features in order to detect covid-19. Every model and its performance will be evaluated with the help of 5fold cross-validation. We attempted to develop a model that can predict whether a person is Covid-19 is positive or negative based on symptoms. Also, to ensure that no positive individual is labelled as Covid affected (minimizing false positive). Implementing a machine learning model to properly, consistently, and efficiently categorize the records in the data collection. Using hyper parameter tweaking to select the ideal parameters for algorithms.

Keywords: COVID-19, prediction, comparative analysis, Python, symptoms.

I. INTRODUCTION

At the end of 2019, the first case of pneumonia of origin was identified in Wuhan, China. Then they identified it as severe acute respiratory syndrome beta coronavirus and novel coronavirus disease. In April 2020, the spread of COVID-19 impacted more than 200 countries.

World health organisation's has given advice for the public to frequently handwashing, disinfecting, social distance, and wearing masks can protect one from being infected. WHO has listed many symptoms ,and they emphasized that fever, cough, and tiredness were common symptoms. And the less common symptoms were headaches: Diarrhea, loss of smell ,etc. [1].

Although there was prediction done by many factors Robotics and AI helped predict and give treatment for covid-19. Robotics and artificial intelligence have an increased presence to help follow the spread of COVID-19 in airports, hospitals, transportation etc. Humanoid robots, autonomous vehicle, and drones were used to reduce human contact and spread the coronavirus. They are used in disinfecting the public spaces, sensing body temperature and making comfort to the travellers.

Robots have served humans by protecting them from any dangerous work. COVID-19 has brought long several threats and restrictions to society. Many potential implementations have been proposed for using robots in healthcare for facing these challenges. Healthcare workers were at the battle's forefront during pandemic times and professional groups with high vulnerability. Robotics and artificial intelligence are two different fields that will be combined for giving many applications in many areas. Robotics will be involved in robot creation for doing many works without any further interaction, whereas AI is how the system emulates the human mind for making decisions and learning. We can robotics with AI elements and AI elements with robotics [2]. Most robots are designed for performing simple and easy work. And there will be no need for AI to be implemented. AI free robotic systems are created with the limitations of AI in mind. In many fields, AI and robotics are used and they are aerospace, healthcare, agriculture, education etc. Healthcare facilities play an important role in handling all pandemics. Robotic applications are useful in such situations where they replicate human actions in unsafe situations where human contact will be reduced.

Apart from robots with the help of the Weka tool, the COVID-19 can be predicted. With many symptoms, the prediction of covid-19 is done in this project. With some common symptoms like cough, fever, loss of smell, taste etc. the prediction is done in this project. Algorithms will be compared and the best algorithm will be selected. From this project, the medical professionals can be benefitted by easily predicting covid-19 with all common symptoms of covid.

II. LITERATURE REVIEW

[3] this article provides befit review of commonly used protocols in the health care system and provides for improving the efficiency of the current system. This article provides a case study where the real story is taken. And for that case study, the case review is made. discussion from the case study is covered in this article. This paper is useful for this project because this article covers home health monitors and personalised medical robots, triaging in emergency rooms with AI technology. Applications of AI and robotics in current use. This paper covers the implementation of some dramatic measures for overcoming the pandemic. AI, robotics and telemedicine give a unique platform for the changes that happen to any device.

[4], aims for highlighting the emerging role of robotic applications in healthcare and allied areas. In this article, a review was conducted based on

various robots that have been implemented during the COVID-19 pandemic for attenuating and containing viruses. The result obtained from this study shows that the implementation of robotics in healthcare fields will control the spread of the COVID-19 pandemic because this blocks the coronavirus propagation that happens between the patients and healthcare workers with advantages. There are many types of medical robots which are discussed in this paper. They include disinfecting or spraying robots which are used as portable robots to clean ad disinfect the objects which are increasing fast. Hospitality robots for playing the role of receptionist and nursing robots are increasing because of a pandemic. There are three types of hospitality robots and they are medical servers, receptionists and nurse robots. Teleoperation and telepresence system are also used as medical robots. Some of the telepresence robots that were implemented after the pandemic were NIGA-BOT, Maitri and Zorabot. The next three medical robots are surgical robots, radiologist robots and rehabilitation robots. surgical robots are used for applying autonomy to surgery and this provides many advantages like stability, and the ability to work in any situation. radiologist robots are used for doing medical imaging for diagnosing patients. rehabilitation robots are used for serving the purpose of a nurse who is injured or not able to get back to a patient's normal condition. [5], analysed three robotics and artificial intelligence for preventing COVID-19 in airports. First is the body temperature measurement function. Measuring the outside temperature has increased with advanced technology while adding identification approaches, gives results with the respect for the patient distance, improves accuracy and creates better monitoring tools. Digital temperature sensors are used here for supporting the wide range of accuracy and performance conditions in many applications. The second is the detection of personnel wearing masks. For preventing people from getting coronavirus, infectious pathogens and blocking larger particles covering face masks are made mandatory in many places. Here advanced LLE-CNN related facia recognition solution is used for identifying and tracking everyone in the crowd for recognising the unmasks people and access restrictions for them. The next one is an intelligent analysis where the fast data will be collected and real-time processing will be done. the data which are returned will be processed automatically which will be useful for making a quick decision.

III. MATERIALS AND METHODS

Data collection

During the process of data collection we came across Israel covid dataset which is suitable for this project, so we used the public data set.

Data processing

Data preparation comprises cleaning, transformation, outlier identification, normalization, and feature engineering

Data modelling

This entails partitioning the dataset into train and test data and using the train data to train the model and the test data to test the trained model's performance.

Target prediction

This is also known as model use, and it involves testing a trained model on fresh, unseen real-world data to determine the model's efficacy.

IV. EXPERIMENTS AND OBSERVATIONS

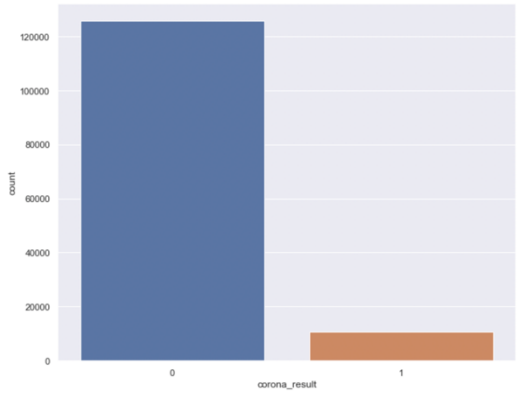
Dataset

```
In [4]: data_set.head(5)
```

Out[4]:

	test_date	cough	fever	sore_throat	shortness_of_breath	head_ache	corona_result	age_60_and_above	gender	test_indication
0	2020-04-30	0	0	0	0	0	negative	None	female	Other
1	2020-04-30	1	0	0	0	0	negative	None	female	Other
2	2020-04-30	0	1	0	0	0	negative	None	male	Other
3	2020-04-30	1	0	0	0	0	negative	None	female	Other
4	2020-04-30	1	0	0	0	0	negative	None	male	Other
5	2020-04-30	1	0	0	0	0	negative	None	female	Other

corona_tested_individuals.csv file is fetched which is downloaded.

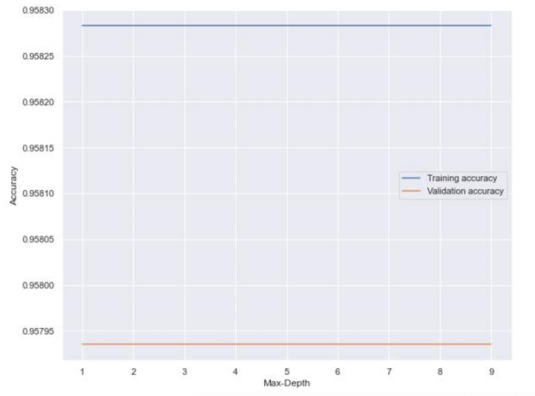


Count plot representing number of positive and negative records.

Decision tree (J48)

One kind of decision tree algorithm is J48 DT which is one of the common and simple decision tree algorithms that are used for classification purposes. It utilises a divide and conquers approach that divides the instances for sub-ranges related to the attribute values [9].

III

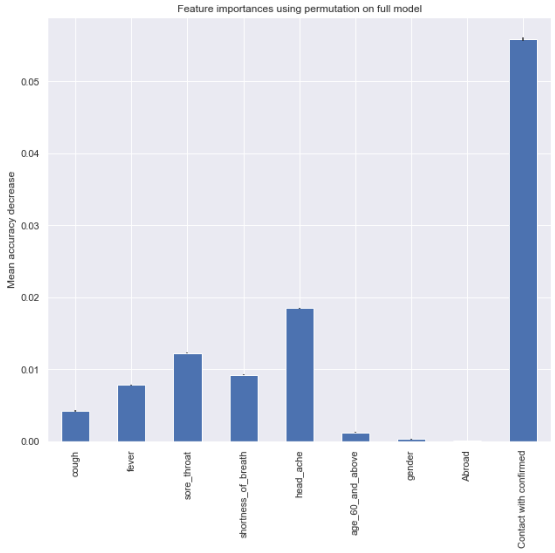


Correctly classified instances – 95.8%
Incorrectly classified instances – 4.2%
R2 score – 0.414
Time taken – 1.08s

While using the J48 algorithm, the accuracy obtained is 95.8 percent. So, medical professionals or doctors can utilise this algorithm for predicting covid-19 easily. The result that the patient has covid or not will be identified easily. and at the same time, the accuracy is 95.8 percent.

Random forest feature importance by random forest

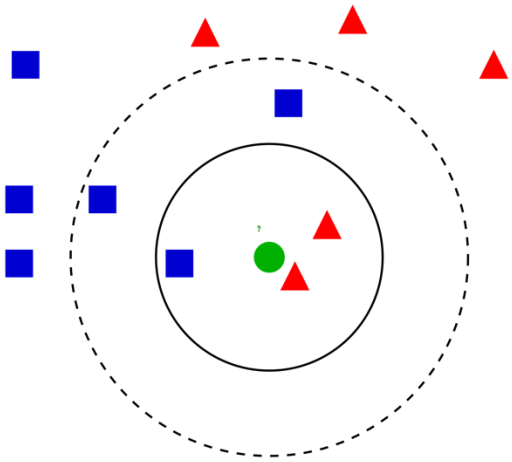
The random forest algorithm produces a tree but this has many trees which are created from the random sample values in the dataset. the final result will be based on the result of the majority of the developed trees [10].



Observation reveals that traits such as Contact With Confirmed, Headache, and Sore Throat are more relevant than others.

KNN

KNN is a type of classification where the function is only approximated locally and all computation is deferred until function evaluation.



Correctly classified instances – 95.09%
Incorrectly classified instances – 4.91%
R2 score – 0.383
Time taken – 83.78s

Logistic Regression

Logistic regression is a process of modeling the probability of a discrete outcome given an input variable. The most common logistic regression models a binary outcome; something that can take two values such as true/false, yes/no, and so on.

Correctly classified instances – 95.06%
Incorrectly classified instances – 4.94%
R2 score – 0.33
Time taken – 7.75s

SVM

In machine learning, support-vector machines are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis.

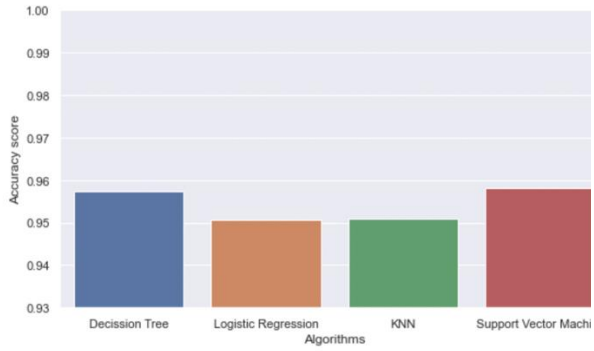
Correctly classified instances – 95.80%
Incorrectly classified instances – 4.20%
R2 score – 0.41
Time taken – 479s

IV

TABLE 1: COMPARATIVE STUDY OF ALL ALGORITHMS EVALUATED

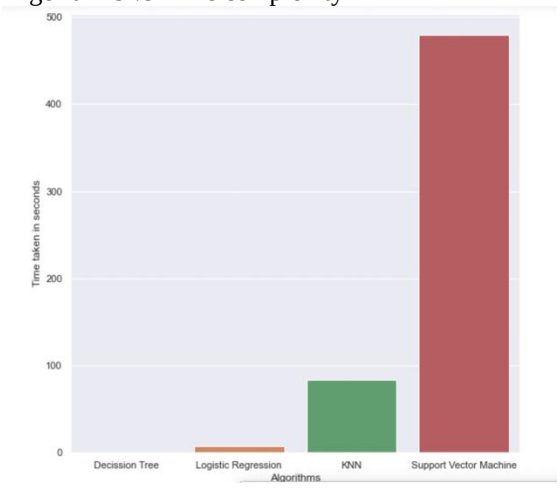
Algorithm	Correctly classified	Incorrectly classified	R2	Time taken
J48	95.8%	4.2%	0.414	1.08s
KNN	95.09%	4.91%	0.383	83.78s
Logistic Regression	95.06%	4.94%	0.33	7.75
SVM	95.80%	4.20%	0.41	479s

Algorithms vs Accuracy Scores

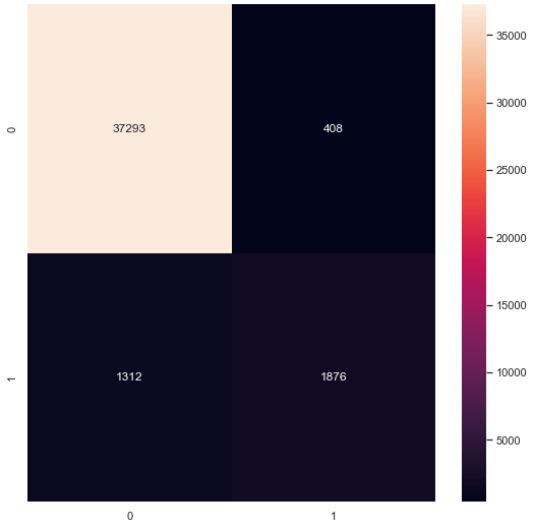


From the above bar plot, which is plotted in between algorithms and accuracy score it is clear that decision tree and support vector machine has performed effectively. Upon observing keenly by considering time complexity decision tree out performs rest of the algorithms.

Algorithms vs Time complexity



Confusion Matrix



{0,0} False Negative
{0,1} False Positive
{1,0} True Negative
{1,1} True Positive

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN}$$

The accuracy, in this case, is around 95%, around 3.2% records who have Covid-19 are classified as healthy. By this example what we are trying to say is that accuracy is not a good metric when the data set is unbalanced. Using accuracy in such scenarios can result in misleading interpretation of results.

V. CONCLUSION

The final aim is to develop a model that employs the most effective machine learning approach for forecasting COVID-19. Also our goal is to give efficient and trustworthy prediction with fewer characteristics and testing. In this study, just nine crucial features are taken into account. Four distinct classification techniques were utilized. Some of the approaches employed include support vector machine, K closet neighbors, logistic regression and decision tree. Before being used in this model, the data was pre-processed. The support vector machine and decision tree are the strategies that give the best results in this paradigm. However , decision trees were explored since they have a reduced time complexity. To widen this, other methodologies like clustering, association rules and genetic algorithms might be applied. Given the limits of this study, there is a need to develop a more complicated and combination of models to

V

achieve greater accuracy for COVID-19 early prediction.

VI. REFERENCES

[1] S. L. S, A. H. S. and V. c. S. S., "Nature inspired optimization model for classification and severity prediction in COVID-19 clinical dataset," *Journal of Ambient Intelligence and Humanized Computing* , pp. 1-13, 2021.

[2] R. Sardar, A. Sharma1 and D. Gupta1*, "Machine Learning Assisted Prediction of Prognostic Biomarkers Associated With COVID-19, Using Clinical and Proteomics Data," *Systems Biology and Omics Approaches to Understand Complex Diseases Biology*, pp. 1-10, 2021.

[3] S. N. Ahmed, "Covid, AI, and Robotics - A Neurologist's Perspective," *Front. Robot. AI*, pp. 1-4, 2021.

[4] S. Raje, N. Reddy, H. Jerbi, P. Randhawa, G. Tsaramirsis, N. V. Shrivas, A. Pavlopoulou, M. Stojmenović and D. Piromalis, "Applications of Healthcare Robots in Combating the COVID-19 Pandemic," *Applied Bionics and Biomechanics*, pp. 1-9, 2021.

[5] S. SENHAJI, S. FAQUIR and M. O. Jamil, "Towards robotics and artificial intelligence for the prevention of Covid 19 pandemic," *E3S Web of Conferences* , pp. 1-8, 2021.

[6] D. M. Matta and M. K. Saraf, "Prediction of COVID-19 using Machine Learning Techniques," *Bachelor of Science in Computer Science*, pp. 1-20, 2020.

[7] C. N. Villavicencio, J. J. E. Macrohon, X. A. Inbaraj, J.-H. Jeng and J.-G. Hsieh, "COVID-19 Prediction Applying Supervised Machine Learning Algorithms with Comparative Analysis Using WEKA," *Algorithms*, pp. 1-22, 2021.

[8] F. E. Shamout, Y. Shen, N. Wu, A. Kaku, J. Park, T. Makino, S. Jastrzębski, J. Witowski, B. Z. Duo Wang, S. Dogra, M. Cao, N. Razavian and D. Kudlowitz, "An artificial intelligence system for predicting the deterioration of COVID-19 patients in the emergency department," *npj Digital Medicine volume* , pp. 1-11, 2021.

[9] D. Painuli, D. Mishra, S. Bhardwaj and M. Aggarwal1, "Forecast and prediction of COVID-19 using machine learning," *Data Science for COVID-19.* , p. 381–397, 2021.

[10] L. Wang, Y. Zhang, D. Wang, X. Tong, T. Liu, S. Zhang, J. Huang, L. Zhang, L. Chen, H. Fan and M. Clarke, "Artificial Intelligence for COVID-19: A Systematic Review," *Medicine*, pp. 1-9, 2021.

[11] L. R. Kolozsvári, T. Bérczes, A. Hajdu, R. Gesztelyi, A. Tiba, I. Varga, A. B.Al-Tammemi, G. J. Szőllősi, S. Harsányi, S. Garbóczy and J. Zsuga, "Predicting the epidemic curve of the coronavirus (SARS-CoV-2) disease (COVID-19) using artificial intelligence: An application on the first and second waves," *Informatics in Medicine Unlocked*, pp. 1-6, 2021.

[12] P. Guleria, S. Ahmed, A. Alhumam and P. N. Srinivasu, "Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States," *Healthcare (Basel).*, pp. 1-8, 2022.

[13]