Feature Learning and Analysis of Pre Existing Conditions Prone to Covid Virus During Second Wave

Dr.Supriya[1], Shami K[2], Sharmistha C[3], Sowjanya V[4], Aneja P[5], Ms.Krathiyayini J[6]

[1]Department of ISE New Horizon College of Engineering Banglore

[2,3,4,5]New Horizon College of Engineering Banglore

[6]Senior Assistant professor, New Horizon College of Engineering Banglore

Abstract - **The major concern for Covid is prevention of spread. It can only be done by fast testing and isolation. But, the fastest test takes two hours by which majority of the spread takes place. We have made a website which takes a dataset from hospital and predicts if the patient is covid positive or not. The model uses the vast Covid-19 test dataset which has various vitals and symptoms along with the patient's history of pre-existing medical conditions and Covid-19 test results as its features. We have trained the dataset using random forest classifier algorithm as it has the maximum accuracy in predicting the patient's final result. This trained model is then used for predicting if the patient is potentially Covid positive or negative based on the input provided from the user. For this purpose we have designed UI as follows. We are using the technologies like Python, Flask framework for predicting the results, html and css code for UI integration. The objective of the added feature is to build a user-friendly application for hospitals which can be used to predict the Covid test results of the patients based on their symptoms and recorded vitals. This way the hospitals authorities can separate the potential Covid patients from the healthy crowd and stop them from infecting healthy individuals before the confirmed reports are released.**

Index Terms – **Introduction, Problem Definition, Design, Procedure, Backend Program, Future Work, Conclusion, Reference.**

I.INTRODUCTION

The 2019 novel Covid illness (COVID19) pandemic brought about by SARSCoV2 keeps on representing a basic and dire danger to worldwide wellbeing. The flare-up in Hubei Province of the People's Republic of China toward the beginning of December 2019 has spread across the world. The complete number of affirmed patients with the illness has surpassed 39,500,000 in> 180 nations, albeit the quantity of those tainted is possible a lot higher. In excess of 1,110,000 individuals have passed on from COVID19. This pandemic keeps on testing clinical frameworks all throughout the planet from various perspectives., remembering a flood for the interest for medical clinic beds and a basic deficiency of clinical gear as numerous medical care laborers become tainted, bringing about prompt clinical choices being made and clinical assets being utilized adequately.

II. PROBLEM DEFINITION

The Covid Predictior Using Random Forest Classifier. Our venture utilizes the huge Covid-19 test dataset which has different vitals and manifestations alongside the patient's set of experiences of prior ailments and Covid-19 test results as its highlights.

We have prepared the dataset utilizing irregular backwoods classifier calculation as it has the most extreme exactness in foreseeing the patient's end-product. This prepared model is then utilized for foreseeing if the patient is possibly Covid positive or negative dependent on the info gave from the client.
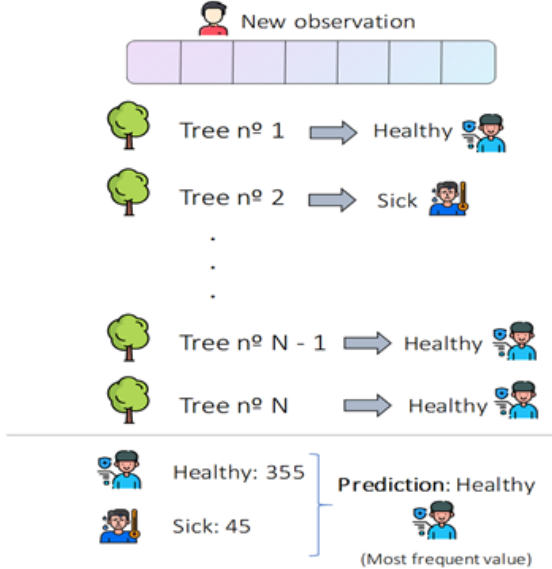
III. DESIGN

Training and Building a Random Forest

Irregular Forest models consolidate the effortlessness of Decision Trees with the adaptability and force of a gathering model. In a woodland of trees, we disregard the high difference of a particular tree, and are less worried about every individual component, so we can develop more pleasant, bigger trees that have more prescient force than a pruned one.

Albeit Random Forest models don't offer as much decipher capacity as a solitary tree, their exhibition is significantly better, and we don't need to stress such a

great amount over impeccably tuning the boundaries of the woods as we do with singular trees.
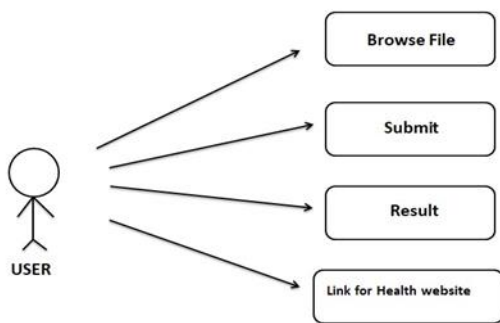

FUNCTIONAL REQUIREMENTS:

Choose File Functionality: The user should upload the csv data file.

Submit Functionality: Once all file is uploaded, we can submit to get the prediction results.

Domain: Deep Learning

UI Design Frontend: HTML, CSS, FLASK (Framework)

USE CASE DIAGRAM

IV. PROCEDURE

1.Browse File

We are using the CSV file for prediction of covid-19 test results, we are taking the different types of test results, tests are either positive or negative and we are also taking the different information like test names,

swab types to upload csv file we need to follow some instructions like we need to check the required format of the columns before uploading.

2.Submit

When we upload the required csv file and click on the submit button it takes actual csv file and produce the predicted covid-19 test results and it also gives the information about the test name, swab type, covid-19 test results.

Main website page to upload and submit CSV file
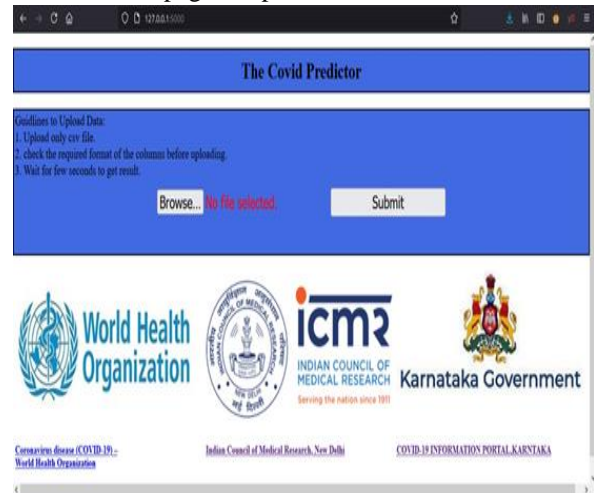
3.Result

Once the CSV file chosen is submitted, we receive the result page which contains the test name conducted on the patient, swab type, the actual covid test results and predicted covid test results.

IV. BACKEND PROGRAM

The model uses the vast Covid-19 test dataset which has various vitals and symptoms along with the patient's history of pre-existing medical conditions and Covid-19 test results as its features. We have trained the dataset using random forest classifier algorithm as it has the maximum accuracy in predicting the patient's final result. This trained model is then used for predicting if the patient is potentially Covid positive or negative based on the input provided from the user.

Accuracy: 0.9857223370279917

Predicting if the patient is covid positive or negative using Random Forest classifier and accuracy prediction

V. FUTURE WORK

As a future work we will propose hybrid model makes it possible to accumulate novel features of the info because it predicts Covid-19 cases supported different scenarios like location, past data, movement patterns of citizens, and so forth. The proposed cross breed administered model can employee a combination of outfit learning strategies, autoregressive, and moving backward models. Using this mixture of techniques helped to simply combine multiple features and inputs, predict future data using past data trends, and produce average aggregate results. The proposed model will most likely have 80%, 10%, and 10% of the overall data for training, testing, and validation, respectively.

For implementation, two different datasets are often taken, the ministry of India dataset and therefore the World meter dataset from the months of February to July 2020. Using this hybrid model, different parameters can be predicted at an international and national level such as the number of affected cases, confirmed cases, and deaths. Particularly in India, the quantity of dynamic cases and passing can be anticipated at a state wise granularity. The fundamental oddity of the proposed work lies in its drawn-out forecast precision, as against different techniques which turn out only for a short term. Utilizing these expectations, we will handily gauge the more drawn-out term direction of Covid-19 cases and use it in choice and legislative issues measures. The future work should join various models like intermittent neural organizations and furthermore utilize nonlinear techniques for forecasts customized to explicit topographical districts.

VI. CONCLUSION

The COVID-19 pandemic outbreak has devastated the entire world and cause a state of worldwide health emergency. Several efforts are performed to combat this pandemic. In this study, we aimed to explore the impact of important signs, chronic disease, preliminary clinical data, and demographic features to predict the mortality and survival of the COVID-19 patients using supervised machine learning algorithms. SMOTE technique was wont to alleviate the info imbalance. The results showed that random forest outperformed the opposite models using 10-fold cross-validation.

The study achieved the accuracy of 0.953 and AUC of 0.999. Despite the many outcome achieved from this proposed model, there's still a requirement for improvement. The models got to be validated using multiple datasets. Furthermore, within the future, we'll incorporate and explore the impact of other clinical features and laboratory results that were identified as significant within the previous studies.

REFERENCES

[1] https://onlinelibrary.wiley.com/doi/10.1002/spe. 2969
[2] https://www.mdpi.com/1424-8220/21/10/3322/pdf
[3] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC 8138040/
[4] https://scikitlearn.org/stable/modules/generated/s klearn.ensemble.RandomForestClassifier.html
[5] https://towardsdatascience.com/understanding-random-forest-58381e0602d2
[6] https://www.tutorialspoint.com/flask/index.htm
[7] https://www.w3schools.com/js/
[8] https://www.w3schools.com/css/css_intro.asp