IMPLEMENTATION TOWARDS SEVERITY PREDICTION OF COVID 19 USING MACHINE LEARNING

[1] Mr. Shinde Sachin D., [2] Prof. Umesh Pawar

[1]PG Student, [2] Assistant Professor
Department of Computer Engineering
S.N.D.COE and RC, Yeola

Abstract: Several outbreak prediction models for COVID-19 are being used by officials around the world to make informed-decisions and enforce relevant control measures. Among the standard models for COVID-19 global pandemic prediction, simple epidemiological and statistical models have received more attention by authorities, and they are popular in the media. Due to a high level of uncertainty and lack of essential data, standard models have shown low accuracy for long-term prediction. Although the literature includes several attempts to address this issue, the essential generalization and robustness abilities of existing models need to be improved. Integration of artificial intelligence (AI) techniques in wireless infrastructure, real-time collection, and processing of end-user devices is now in high demand. It is now superlative to use AI to detect and predict pandemics of a colossal nature. The Coronavirus disease 2019 (COVID-19) pandemic, which originated in Wuhan China, has had disastrous effects on the global community and has overburdened advanced healthcare systems throughout the world. Globally; over 4,063,525 confirmed cases and 282,244 deaths have been recorded as of 11th May 2020, according to the European Centre for Disease Prevention and Control agency. However, the current rapid and exponential rise in the number of patients has necessitated efficient and quick prediction of the possible outcome of an infected patient for appropriate treatment using AI techniques. This paper proposes a fine-tuned SVM model boosted by algorithm. The model uses the COVID-19 patient's age, various symptoms such as dry cough, fever, etc and it will compare with the existing dataset to predict the severity of Covid 19.

***Index Terms*— Covid 19, severity, prediction, python, svm, etc**

I. INTRODUCTION

The catastrophic outbreak of severe acute respiratory syndrome - Coronavirus (SARS-CoV-2) also known as COVID-2019 has brought the worldwide threat to the living society. The whole world is putting incredible efforts to fight against the spread of this deadly disease in terms of infrastructure, finance, data sources, protective gears, life-risk treatments and several other resources. The artificial intelligence researchers are focusing their expertise knowledge to develop mathematical models for analyzing this epidemic situation using nationwide shared data. To contribute towards the well-being of living society, this article proposes to utilize the machine learning models with the aim for understanding its severity prediction comparing user dataset and standard dataset. Multiple cases of pneumonia patients were linked to the coronavirus disease-2019 (COVID-19) occurred in December 2019 (Zhu et al., 2020). The virus 2019-nCoV demonstrated a substantial capability of inter-human transmissions (Chan et al., 2020) and has rapidly spread around the world, in particular South Korea and Japan (Li Q. et al., 2020). Patients infected with COVID-19 had significantly varied symptoms and their outcomes ranged from mild to death, and the mortality rate was approximately 4.3\% (Wang et al., 2020). It is necessary to mention that 61.5\% of the COVID-19 pneumonia patients with critical symptoms died within 28 days after admission (Yang X. et al., 2020). The discrimination of severely ill patients with COVID-19 from those with mild symptoms may help understand the individualized variations of the COVID-19 prognosis. The knowledge may also facilitate the establishing of early diagnosis of the COVID-19 severeness. The diagnosis of COVID-19 heavily relies on the epidemiological features, clinical characteristics, imaging findings, and nucleic acid screening (Shi et al., 2020), etc. The delivery of the diagnosis result by these technologies was time consuming and error prone (Xie et al., 2020). Multiple types of clinical data were collected for a patient with COVID-19 infection and they were manually integrated by the clinicians to make the diagnosis decisions. The stochastic transmission model was also used to investigate how the COVID-19 transmitted locally and globally (Kucharski et al., 2020). Machine learning algorithms were widely used to integrate the heterogeneous biomedical data sources for the diagnosis decision (Thompson et al., 2018; Hu et al., 2019). So they may also be utilized to produce more delicate prediction models for the severeness diagnosis of the COVID-19 patients. The biomarkers used for an accurate diagnosis model of patients with COVID-19 may serve as the drug targets for this global infectious disease. This study investigated the detection of severely ill patients with COVID-19 from those with mild symptoms using the clinical information and the blood/urine test data. The clinical information consisted of age, dry cough, difficulty in breathing, Sore-Throat, pains, Nasal-Congestion, diarrhea, etc. An accurate severeness detection model of the patients with COVID-19 based on those features above may improve the prognosis of this disease in large scale clinical practices. The following sections will firstly describe the data collection and modeling methods, and then utilized the popular machine learning algorithms to build the best severeness detection model.

Paper is organized as follows. Section II describes about the related work done earlier for the system to be developed. Section III presents method used and algorithms used for the detection. Section IV presents experimental results showing results of images tested. Finally, Section V presents conclusion.

II. RELATED WORK

1) Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts

This paper aims at emphasizing their importance in responding to the COVID-19 outbreak and preventing the severe effects of the COVID-19 pandemic. We firstly present an overview of AI and big data, then identify the applications aimed at fighting against COVID-19, next highlight challenges and issues associated with state-of-the-art solutions, and finally come up with recommendations for the communications to effectively control the COVID-19 situation. It is expected that this paper provides researchers and communities with new insights into the ways AI and big data improve the COVID-19 situation, and drives further studies in stopping the COVID-19 outbreak.

2) Application of Machine Learning in Disease Prediction

Pahulpreet Singh Kohli 2018

In this paper, we apply different classification algorithms, each with its own advantage on three separate databases of disease (Heart, Breast cancer, Diabetes) available in UCI repository for disease prediction. The feature selection for each dataset was accomplished by backward modeling using the p-value test. The results of the study strengthen the idea of the application of machine learning in early detection of diseases.

3) Construction of TCM Health Management Model for Patients with Convalescence of Coronavirus Disease Based on Artificial Intelligence Fan GAO 2020

This article will tap the advantages of TCM resources, improve the quality of patient recovery "Rehabilitation", break the time and space limitations of health management and provide TCM health management methods for patients with convalescence with a view to human health promotion.

4) COVID-19 Optimizer Algorithm, Modeling and Controlling of Coronavirus Distribution Process

This paper first proposes a novel COVID- 19 optimizer Algorithm (CVA) to cover almost all feasible regions of the optimization problems. We also simulate the coronavirus distribution process in several countries around the globe. Then, we model a coronavirus distribution process as an optimization problem to minimize the number of COVID-19 infected countries and hence slow down the epidemic spread. Furthermore, we propose three scenarios to solve the optimization problem using most effective factors in the distribution process. Simulation results show one of the controlling scenarios outperform the others. Extensive simulations using several optimization problems show that the CVA technique performs best with up to 15%, 37%, 53% and 59% increase compared with Volcano Eruption Algorithm (VEA), Gray Wolf Optimizer (GWO), Particle Swarm Optimization (PSO) and Genetic Algorithm (GA), respectively.

5) Designing Disease Prediction Model Using Machine Learning Approach Dhiraj Dahiwade, 2Prof. Gajanan Patle, 2Prof. Ektaa Meshram

The proposed general disease prediction based on symptoms of the patient. For the disease prediction, we use K-Nearest Neighbor (KNN) and Convolutional neural network (CNN) machine learning algorithm for accurate prediction of disease. For disease prediction required disease symptoms dataset. In this general disease prediction the living habits of person and checkup information consider for the accurate prediction. The accuracy of general disease prediction by using CNN is 84.5% which is more than KNN algorithm. And the time and the memory requirement is also more in KNN than CNN. After general disease prediction, this system able to gives the risk associated with general disease which is lower risk of general disease or higher.

6) On the Analysis of COVID19 – Novel Corona Viral Disease Pandemic Spread Data Using Machine Learning Techniques Shreyas Setlur Arun

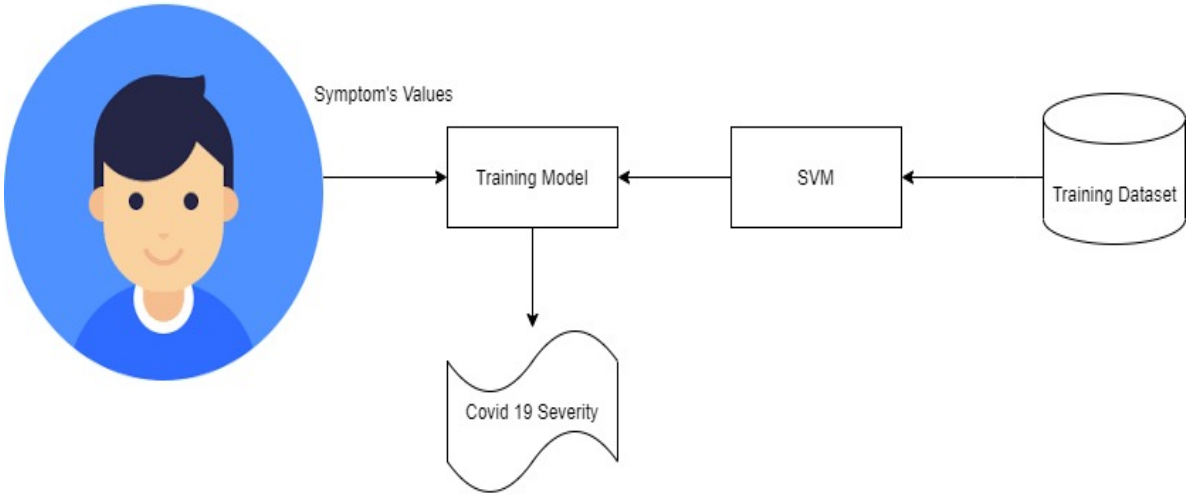
Coronaviruses are a group of viruses that cause various diseases in mammals and birds. In humans, they cause a range of respiratory disorders. This paper presents the analysis of the transmission of COVID19 disease and predicts the scale of the pandemic, the recovery rate as well as the fatality rate. We have used some of the well-known machine learning techniques as well as mathematical modeling techniques such as Rough Set- Support Vector Machine (RS-SVM), Bayesian Ridge and Polynomial Regression, S IR model , and RNN. work.

III. METHODOLOGY

1) METHODOLOGY

User will register and login to the system where user can enter the symptoms, he/she suffering from. The SVM will take standard dataset for training purpose and will create a training model. then user enters the symptoms, it will be compared with the training model and user will get the severity prediction as per the training model and standard dataset. The SVM algorithm will be used for trainingpurpose.

2) SYSTEM ARCHITECTURE



.

3) MODULES

A) Data Pre-processing & Classification

A technique which is used to transform the raw data in a useful and efficient format. In System , symptoms are the Input Data .For pre preprocessing , Cleaning and tokenization id done. Classification is Data mining task of the predicting value of Categorical variable. This is done by building model depending on one or more attributes or features

B) Training And Testing

Training data:

This type of data builds up the machine learning algorithm. The data scientist feeds the algorithm input data, which corresponds to an expected output. The model evaluates the data repeatedly to learn more about the data's behavior and then adjusts itself to serve its intended purpose.

Test data:

After the model is built, testing data once again validates that it can make accurate predictions. If training and validation data include labels to monitor performance metrics of the model, the testing data should be unlabeled. Test data provides a final, real-world check of an unseen dataset to confirm that the ML algorithm was trained effectively.

C)Prediction

Prediction of covid patients is done depending on symptoms data. Prediction for number of patients is done depending on data collected from worldometer.

4) MATHEMATICAL MODEL

Let

S be Closed system defined as, S = Ip, Op, Ss, Su, Fi, A

To select the input from the system and perform various actions from the set of actions A so that Su state can be attained.

S=Ip,Op,Ss,Su,Fi,A

Where,

IP1=Username,Password, Parameters

Set of actions=A=F1,F2,F3,F4

Where

F1= Preprocessing

F2= classification

F3= Analysis

F4= Severity of COVID 19 Detection

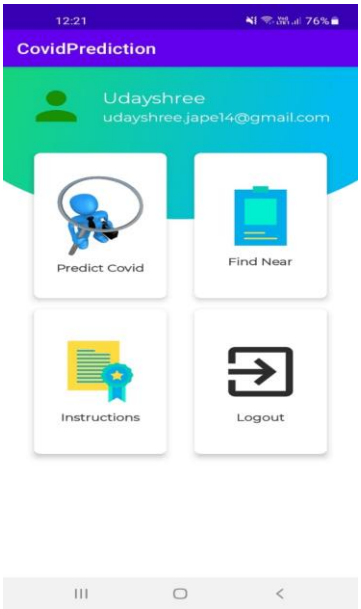
S=Set of users
Ss=rest state, registration state, login state
Su- success state is successful analysis
Fi- failure state.
Objects:
1) Input1: Ip1 = Username, Password
2) Input2 : Ip2= Input Parameters1) Output1 : Op1 = Authentication and Data Processing02) Output2 : Op2 = Classification
3) Output3 : Op3 = severity detection of covid-19

5) Algorithms:
SVM :

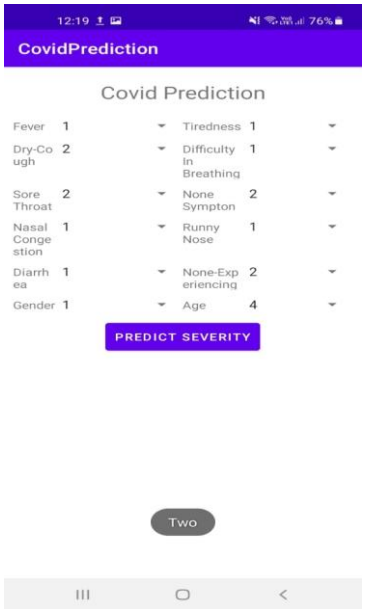
```
Input: D dataset, on-demand features, aggregation-based features,
Output: Classification of Application
for each application App-id in D do
Get on-demand features and stored on vector x for App-id
x.add ( Get-Features(app-id));
end for
for each application in x vector do
Fetch first feature and stored in b, and other features in w.
hw,b (x) = g (z) here z= ( wT x + b)
if (z 0)
assign g(z)=1;
else g(z)=-1;
end if
end for
```

IV. RESULTS AND DISCUSSIONS

A) APPLICATION SETUP



B) ANALYSIS



The purpose of the following study is to accurately predict the outcome of a particular patient depending on multiple factors, including but not limited to travel history, demographics etc. since this is a very crucial prediction, accuracy is very important. thus, for the purpose of evaluating the model we considered three evaluation metrics for this study.
The following terms are used in the equations:
TP, TRUE POSITIVE; TN, TRUE NEGATIVE; FP, FALSE POSITIVE; AND FN, FALSE NEGATIVE.

ACCURACY
Given a dataset consisting of (tp + tn) data points, the accuracy is equal to the ratio of total correct predictions (tp + tn + fp + fn) by the classifier to the total data points. accuracy is an important measure which is used to assess the performance of the classification model. accuracy is calculated as shown in equation (1) as follows:

$$\text{ACCURACY}=\text{TP + TNTP + TN + FP + FN} \qquad 0.0<\text{ACCURACY}<1.0 \quad (1)$$

PRECISION
Precision is equal to the ratio of the true positive (tp) samples to the sum of true positive (tp) and false positive (fp) samples.precision is calculated as given in equation (2) as follows:

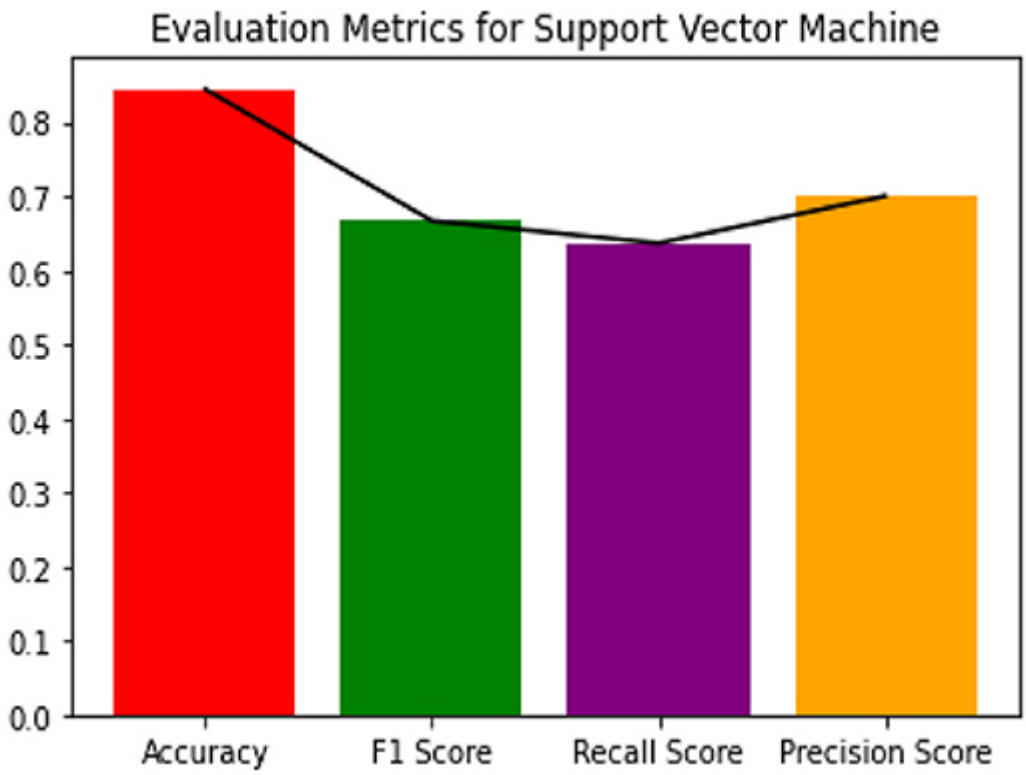
$$\text{PRECISION}=\text{TPTP + FP} \quad (2)$$

RECALL
Recall is equal to the ratio of the true positive (tp) samples to the sum of true positive (tp) and false negative (fn) samples.recall is calculated as given in equation (3) as follows:

$$\text{RECALL}=\text{TPTP + FN} \quad (3)$$

F1 SCORE
F1 score is equal to the harmonic mean of recall and precision value. the f1 score strikes the perfect balance between precision and recall thereby providing a correct evaluation of the model's performance in classifying covid-19atients. this is the most significant measure that we will be using to evaluate the model. f1 score can be calculated as shown in equation (4) as follows:

$$\text{F1 SCORE}=2 \times \text{PRECISION} \times \text{RECALLPRECISION + RECALL} \quad (4)$$



V. CONCLUSION

Thus, we are going to implement a system using python as a programming language for covid 19 severity detection using ML. SVM algorithm will be used for training model purpose. This system will help covid patients upto great extent to detect the severity. The system will also useful for research-based models. In future we will try to get the sponsorship for the project from Government of India, also we can try to gather the data set from different resources which will improve the efficiency of the system. Also, if we can get the sponsorship from the government, we can implement the app with commercial servers and space which will increase the scope of our project.

REFERENCES

1. Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts QUOC-VIET PHAM1, DINH C. NGUYEN2, THIEN HUYNH-THE3, WON-JOO HWANG4,5, AND PUBUDU N. PATHIRANA2 DOI 10.1109/ACCESS.2020.3009328, IEEE Access
2. Construction of TCM Health Management Model for Patients with Convalescence of Coronavirus Disease Based on Artificial Intelligence Fan GAO 2020 International Conference on Big Data and Informatization Education (ICBDIE) DOI 10.1109/ICBDIE50010.2020.00104
3. COVID-19 Optimizer Algorithm, Modeling and Controlling of Coronavirus Distribution Process Eghbal Hosseini, Kayhan Zrar Ghafoor, Ali Safaa Sadiq, Mohsen Guizani, Ali Emrouznejad DOI: 10.1109/JBHI.2020.3012487, IEEE Journal of Biomedical and Health Informatics.
4. On the Analysis of COVID19 – Novel Corona Viral Disease Pandemic Spread Data Using Machine Learning Techniques Shreyas Setlur Arun 978-1-7281-4876-2/20/\$31.00 ©2020 IEEE
5. J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, S. Hu, Y. Wang, X. Hu, B. Zheng et al., "Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study," medRxiv, 2020.
6. M. Holshue, C. DeBolt, and L. S. First, "Novel coronavirus in the united states," N Engl J Med, vol. 2020, p. 31, 2019.
7. H. S. Maghdid, A. T. Asaad, K. Z. Ghafoor, A. S. Sadiq, and M. K. Khan, "Diagnosing covid-19 pneumonia from x-ray and ct images using deep learning and transfer learning algorithms," arXiv preprint arXiv:2004.00038, 2020.
8. "Cnn health," https://edition.cnn.com/2020/03/11/health/coronaviruspandemic- world-health-organization/index.html, Accessed in March, 2020.
9. B. N. Wetsman, "What is contact tracing?" https://www.theverge.com/ 2020/ 4/10/21216550/contact-tracingcoronavirus-what-is-tracking-spread-how-it-works, April,2020.
10. H. S. Maghdid, K. Z. Ghafoor, A. S. Sadiq, K. Curran, and K. Rabie, "A novel ai-enabled framework to diagnose coronavirus covid 19 using smartphone embedded sensors: Design study," arXiv preprint arXiv:2003.07434, 2020.
11. A. Wilder-Smith, C. J. Chiew, and V. J. Lee, "Can we contain the covid- 19 outbreak with the same measures as for sars?" The Lancet Infectious Diseases, 2020.
12. H. Shelani. Number of covid-19 cases reaches 176 in kurdistan region; iraq's total now 703. [Online]. Available: https://www.kurdistan24.net/ en/news/64f6c6f0-6416-471f-b018-47c133ecc8c3note=
13. H. S. Maghdid and K. Z. Ghafoor, "A smartphone enabled approach to manage covid-19 lockdown and economic crisis," arXiv preprint arXiv:2004.12240, 2020.
14. R. M. Anderson and R. M. May, "Population biology of infectious diseases: Part i," Nature, vol. 280, no. 5721, pp. 361–367, 1979.
15. B. C. Malathesh, G. S. Gowda, C. N. Kumar, M. Narayana, and S. B. Math, "Response to: Rethinking online mental health services in china during the covid-19 epidemic," Asian journal of psychiatry, 2020.
16. K. Roosa, Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J. Hyman, P. Yan, and G. Chowell, "Real-time forecasts of the covid-19 epidemic in china from february 5th to february 24th, 2020," Infectious Disease Modelling, vol. 5, pp. 256–263, 2020.
17. Designing Disease Prediction Model Using Machine Learning Approach Dhiraj Dahiwade, 2Prof. Gajanan Patle, 2Prof. Ektaa Meshram 978-1-5386-7808-4/19/\$31.00 ©2019 IEEE
18. Application of Machine Learning in Disease Prediction Pahulpreet Singh Kohli 2018 4th International Conference on Computing Communication and Automation (ICCCA) \%978-1-5386-6947-1/18/\$31.00 \©2018 IEEEpublication.