

# CORONA VIRUS DETECTION USING PROBABILITY CLASSIFICATION

Yasaswini K<sup>1</sup> Sreya M<sup>2</sup> Shirisha B<sup>3</sup> Srilatha A<sup>4</sup> Saicharan CH<sup>5</sup>

<sup>1</sup>Department of CSE, Sreyas Institute of Engineering and Technology  
[yasaswini@sreyas.ac.in](mailto:yasaswini@sreyas.ac.in)<sup>1</sup>

## **Abstract**

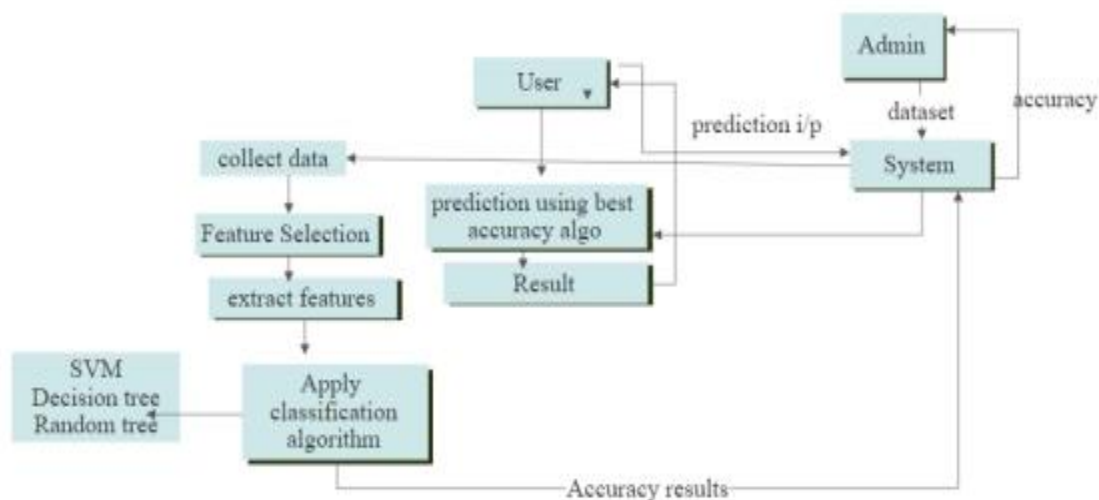
*Covid-19 has become a new problem to the society. It is one of the major epidemic in current days. Everyone in all countries facing this virus. In every country infection as well as the death rate is growing rapidly. Due to this disaster economic status of the world is also decreasing rapidly. To make a break in spreading of virus it is more essential to detect the infected people at an early stage. So to provide a better testing for covid-19 virus detection we are going to develop a project based on considering the previous data of Covid-19 patients. we will be using the Machine learning classification algorithms in our project. So we will be considering the attributes which are common in all the patients so that analyzing the data will be quick. We have collected the 980 patients data and we normalized to 5 attributes such as Temperature, Age, Body Pains, Runny nose and Breathing problem. So by using Random Forest(RF) and Decision Tree(DT) we can train and test the data and we can get result. Among this algorithms we observed that Decision tree gives astonishing performance.*

**Keywords:** Random Forest, Decision Tree, Corona Virus, Machine Learning

## **1. Introduction**

We have selected this project to provide a accurate result of Covid-19 Virus. By developing this project it will overcome some of the problems like People who are hurry for their result they can get result quickly, Poor people families can't afford more money for testing so they can use this one as it is less expensive compared to testing which is happen in hospitals. Covid-19 has become a new problem to the society. Everyone in all Countries are facing this virus .So to provide a better testing for covid-19 virus detection we are going to develop a project based on considering the previous data of Covid-19 patients. There is huge amount of data present in the Data base where hospitals are not using for further Analysis. So by using that data and some Algorithms we are developing a project. As the technology is rapidly increasing we can use many methods and new technologies to overcome problems in medical areas. So by using this new technologies we will get to know more about the disease ,So based on that we can provide many new testings which will help the patients to cure their disease in early stages which will not lead to loss of life. So we will be using the Machine learning classification algorithms in our project. So we will be considering the attributes which are common in all the patients so that analyzing the data will be quick. We have collected the 980 patients data and we normalized to 5 attributes such as Temperature, Age, Body Pains, Runny nose and Breathing problem. So by using Random Forest(RF) and Decision Tree(DT) we can train and test the data and we can get result The main Goal is to create an automated COVID-19 infection probability classification system by using a machine learning technique for early detection of COVID-19.

## How Diagnosis System Works:



**Fig1: Diagnosis System**

## 2. Literature Survey

1. S. K. Wasan, V. Bhatnagar, and H. Kaur, "Impact of data mining techniques on medical diagnostics," *Data Science Journal*, vol. 5, pp. 119-126, 2006.
  - For exploring the hidden patterns in the datasets of the medical domain medical data mining has great potential. These patterns can be utilized for clinical diagnosis.
  - Collection of data Should be done in an organized form. Collected data can be then integrated to form a hospital information system. Data mining technology provides a user-oriented approach to novel and hidden patterns in the data.
2. J. Wu, P. Zhang, L. Zhang, W. Meng, J. Li, C. Tong, "Rapid & accurate identification of COVID-19 infection through machine learning based on clinical available blood test results," *medRxiv*, 2020.
  - Since the sudden outbreak of coronavirus disease 2019, it has rapidly evolved into a momentous global health concern. Due to the lack of constructive information on the pathogenesis of Corona virus and specific treatment, it highlights the importance of early diagnosis and timely treatment.
  - The method presented robust outcome to accurately identify COVID-19 from a variety of suspected patients with similar CT information and symptoms, with accuracy of 0.9795 and 0.9697 for the cross-validation set and test set, respectively.
3. S. Kadry, V. Rajinikanth, S. Rho, N. S. M. Raja, V. S. Rao, and K. P. Thanaraj, "Development of a Machine-Learning System to Classify Lung CT Scan Images into Normal/COVID-19 Class," *arXiv preprint arXiv:2004.13122*, 2020.

- Recently, the lung infection due to Coronavirus Disease (COVID-19) affected a large human group worldwide and assessment of the infection rate in the lung is essential for treatment planning.
  - This research aims to propose a Machine-Learning-System (MLS) to detect the COVID- 19 infection using the CT scan Slices (CTS).
4. T. Singhal, "Review of coronavirus disease-2019 (COVID-19)," The Indian Journal of Pediatrics, pp. 1-6, 2020.
- There is a new public health crises threatening the world with the emergence and spread of 2019 novel coronavirus or the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).
  - The virus originated in bats and was transmitted to humans yet unknown intermediary animals in Wuhan, Hubei province, China in December 2019. There have been around 96,000 reported cases of coronavirus disease and 3300 reported deaths to date (05/03/2020).
  - The disease is transmitted by inhalation or contact with infected droplets and the incubation period ranges from 2 to 14 d.

### 3.Proposed Method

Proposed System consists of automated diagnosis system which is the proper analysis of the data and accuracy. This automated diagnosis system gives rapid results with fewer false negatives. Our proposed system uses machine learning and data mining techniques for extracting only useful information .And classifiers like Decision tree, Random forest, Support vector machine are used among this decision tree shown astonishing performance.

So the technologies that we are going to use are VSCODE(For writing code ),we required a Data base to store all the details so we are using MYsql. In VSCODE we are using python language to write the code. For the frontend development we used PyQt5 which is a Cross platform.

#### Advantages:

- 1.Rapid results
- 2.Fewer false negatives
- 3.Minimally invasive

### 4.Existing Method

Generally in Hospitals the Covid-19 detection is done by testing and using various methods, which is time taking process and in some cases molecular tests have been shown to produce results that say patient doesn't have virus when they actually do. Deep nasal swabs can be uncomfortable for some people especially small children.

#### Disadvantages:

1. Long turnaround times
2. False negatives
3. Uncomfortable for some people

## 5.Implementation

Implementation includes all those activities that take place to convert from old system to new system. The old system takes long turn around time whereas new system gives rapid results

### Admin:

Admin use the dataset of covid patient’s data history for classifications. Admin, train the dataset with following algorithms.

- 1.Decision tree
2. Random Tree.

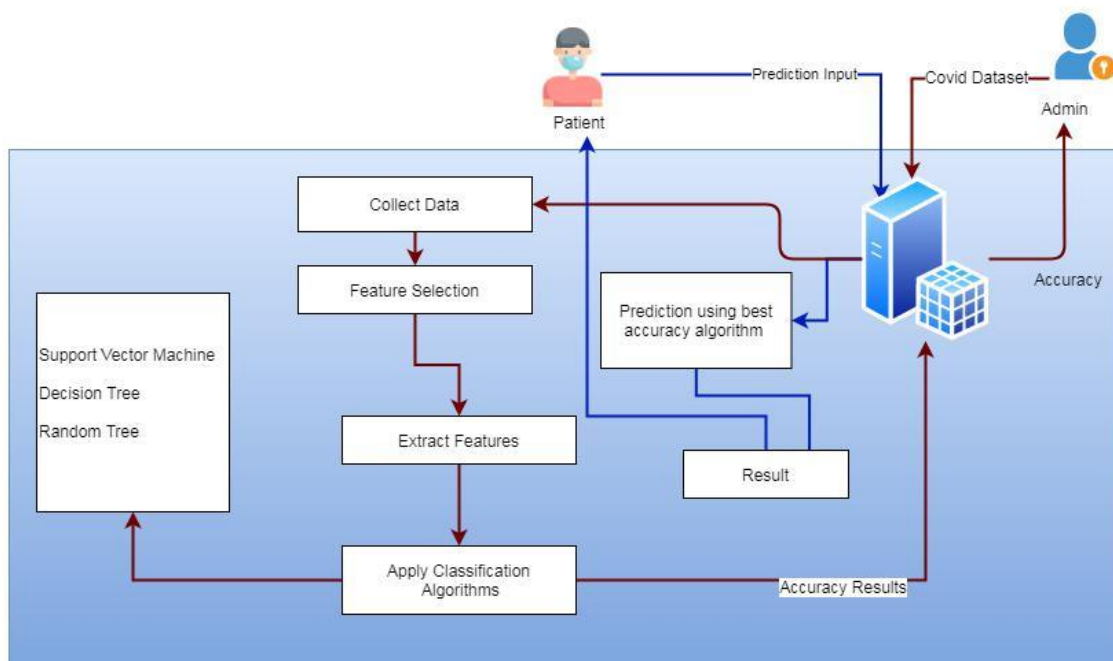
After training the dataset, By using the test data accuracy will be tested by admin. Then the best classification algorithm is founded by admin. Admin also can see the graph of the accuracy of the algorithms and find accuracy scores of all algorithms.

### User:

End user of the application; our application will help to the user by prediction Covid disease by training the previous patient’s dataset with best accuracy algorithm. User can register with own details and after login user can enter details of his/her medical parameter like fever, cough etc. And user can get the result with prediction of best accuracy algorithm.

### System:

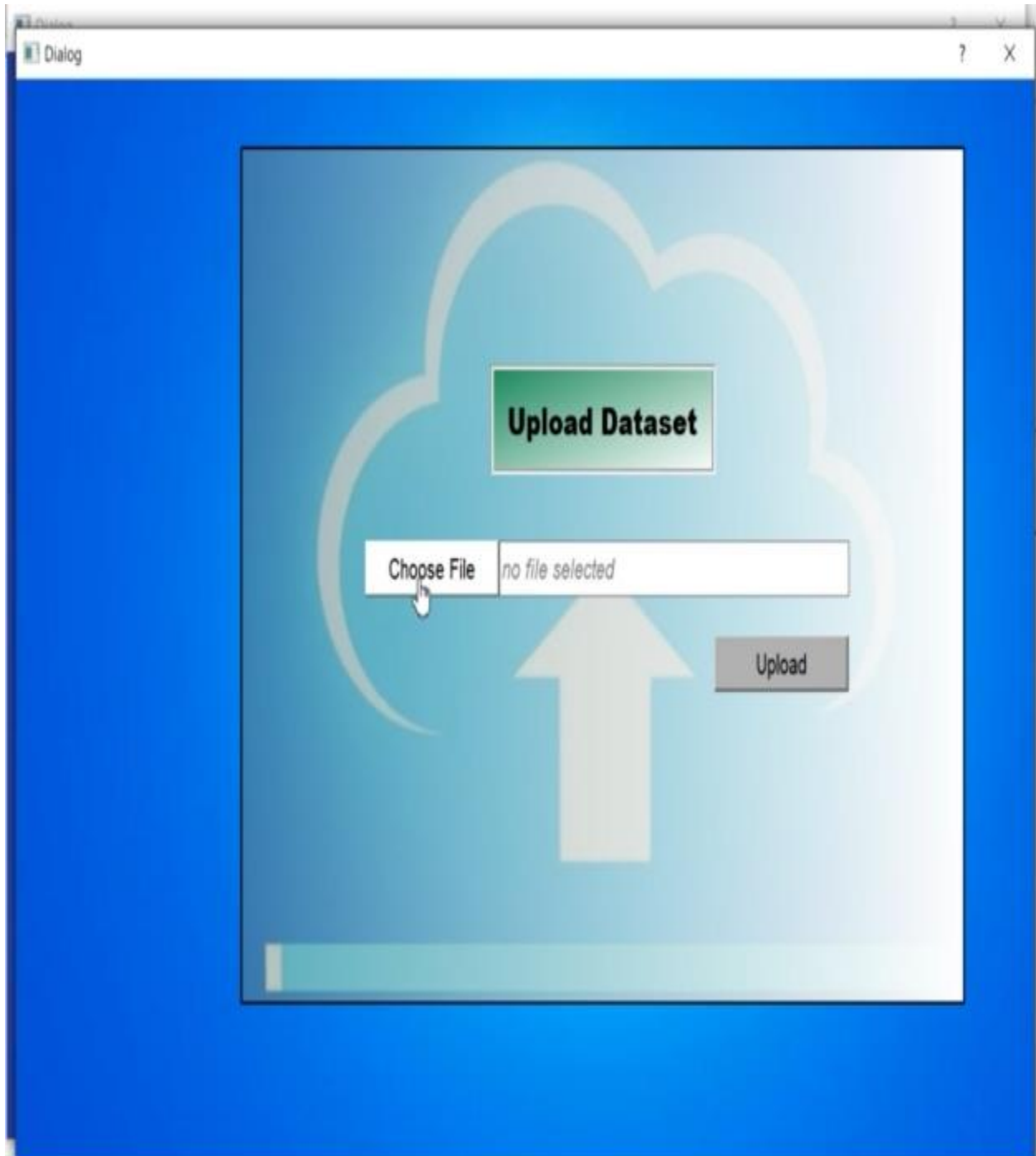
Our system is developed in python with PyQt5 interface components with user friendly. System will interact with database and process the every action of user and admin inputs.



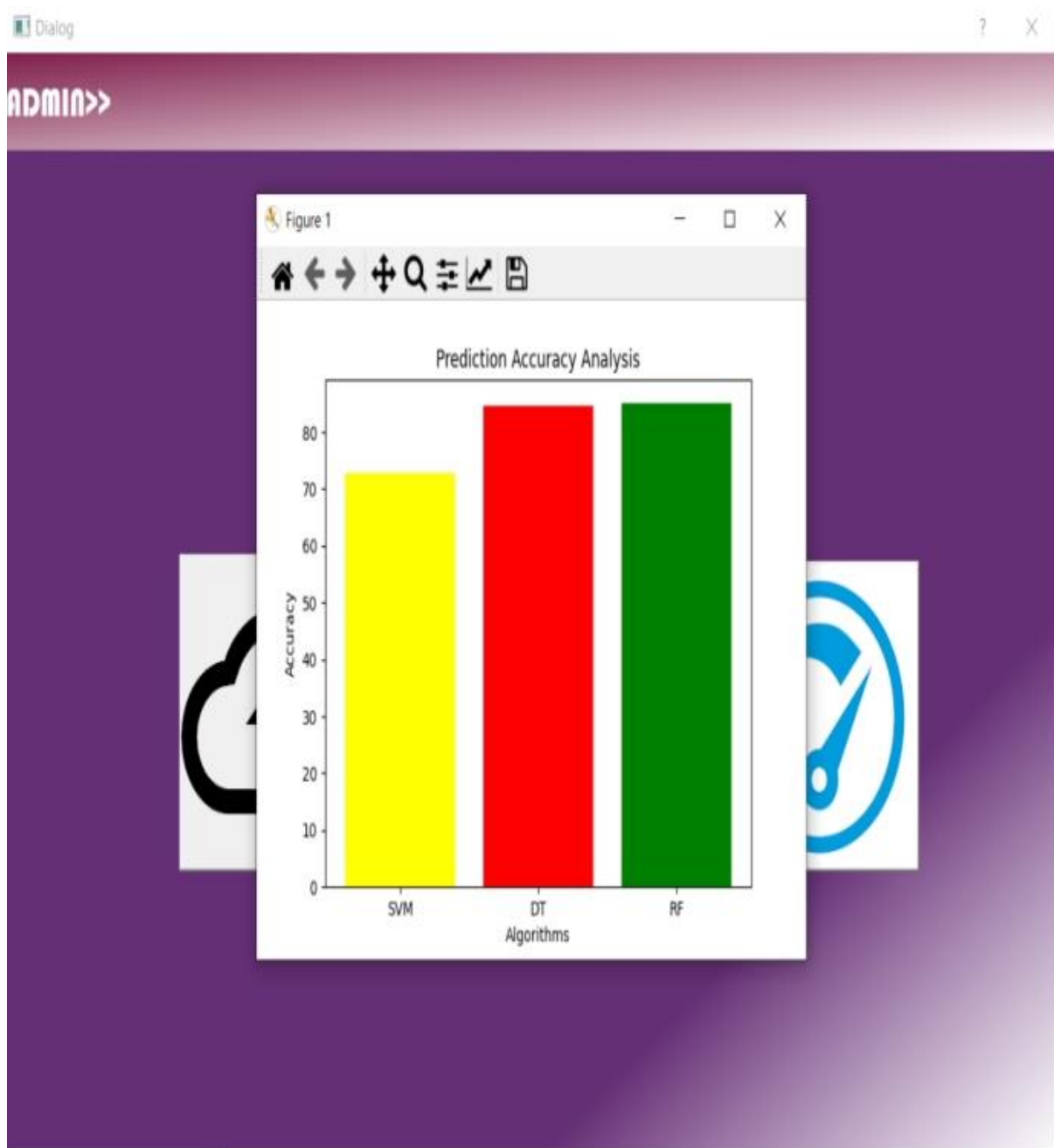
**Fig2: Architecture Diagram**

A system architecture diagram would be used to show the relationship between different components. Usually they are created for systems which include hardware and software and these are represented in the diagram to show the interaction between them. However, it can also be created for web applications.

## 6.Results



**Fig3: Admin Dataset Upload**



**Fig4: Prediction Graph**

## 7. Conclusion

The main objective of our work is to develop an automatic infection probability classification system using ML technique. The model is used for classifying the chance of infection probability of the persons having different COVID-19 symptoms.

- We took a total of 980 samples of data in the dataset and from that, were able to predict covid outcome with good accuracy.
- The performance is measured in terms of accuracy, precision, and recall. Ofthem,

we observed that both Decision tree and Random Forest algorithms work efficiently with almost same performance.

## 8.Future Scope

- Future scope of our project is it will accelerate the processing power of the automation system used in various technologies.
- This project allows for multiple algorithms to be integrated together as modules and results are combined to increase accuracy of final result
- This automated diagnosis system is the proper analysis of the data and accuracy. Numerous machine learning techniques need to use for getting a satisfactory result in various biomedical data analysis.
- We plan to make this a web and mobile application

## 9.References

- [1] S. K. Wasan, V. Bhatnagar, and H. Kaur, "Impact of data mining techniques on medical diagnostics," *Data Science Journal*, vol. 5, pp. 119-126, 2006.
- [2] J. Wu, P. Zhang, L. Zhang, W. Meng, J. Li, C. Tong, et al., "Rapid & accurate identification of COVID-19 infection through machine learning based on clinical available blood test results," *medRxiv*, 2020.
- [3] S. Kadry, V. Rajinikanth, S. Rho, N. S. M. Raja, V. S. Rao, and K. P. Thanaraj, "Development of Machine-Learning System to Classify Lung CT Scan Images into Normal/COVID-19 Class," *arXiv preprint arXiv:2004.13122*, 2020.
- [4] T. Singhal, "A review of coronavirus disease-2019 (COVID-19)," *The Indian Journal of Pediatrics*, pp. 1-6, 2020.
- [5] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, et al., "A novel coronavirus from patients with pneumonia in China, 2019," *New England Journal of Medicine*, 2020.
- [6] T. Bhatnagar, M. V. Murhekar, M. Soneja, N. Gupta, S. Giri, N. Wig, et al., "Lopinavir/ritonavir combination therapy amongst symptomatic coronavirus disease 2019 patients in India: Protocol for restricted public health emergency use," *Indian Journal of Medical Research*, vol. 151, p. 184, 2020.
- [7] C. Cai, L. Han, X. Chen, Z. Cao, and Y. Chen, "Prediction of functional class of the SARS coronavirus proteins by a statistical learning method," *Journal of proteome research*, vol. 4, pp. 1855-1862, 2005.
- [8] A. Rajkumar and G. S. Reena, "Diagnosis of heart disease using datamining algorithm," *Global journal of computer science and technology*, vol. 10, pp. 38-43, 2010.
- [9] L. Chen, H. Liu, W. Liu, J. Liu, K. Liu, J. Shang, et al., "Analysis of clinical features of 29 patients with 2019 novel coronavirus pneumonia," *Zhonghua jie he he hu xi za zhi= Zhonghua jiehe he huxi zazhi= Chinese journal of tuberculosis and respiratory diseases*, vol. 43, pp. E005-E005, 2020.
- [10] D. Orso, N. Federici, R. Copetti, L. Vetrugno, and T. Bove, "Infodemic and the spread Of fake news in the COVID-19-era," *European Journal of Emergency Medicine*, 2020.
- [11] A. B. Kadam and S. R. Atre, "Social media panic and COVID-19 in India," *Journal of Travel Medicine*, 2020.
- [12] S. Haykin, *Neural networks: a comprehensive foundation: Prentice Hall PTR*, 1994.