

# Application of Machine Learning for COVID-19 Data Analysis

Manisha Shinde-Pawar<sup>1</sup>, Rajendra Pujari<sup>2</sup>, Ayesha Mujawar<sup>3</sup>, Alok Shah<sup>4</sup>, Deepali Gala<sup>5</sup>, Bhaskar Patil<sup>6</sup>

<sup>1,2,3</sup>Bharati Vidyapeeth (Deemed to be University), Pune, Institute of Management and Rural Development Administration, Sangli

<sup>4</sup>Bharati Vidyapeeth (Deemed to be University), Pune, Department of Management Studies, Navi Mumbai

<sup>5,6</sup>Bharati Vidyapeeth (Deemed to be University), Pune, Institute of Management, Kolhapur

Email: mjs.imrda@gmail.com

## Abstract

The entire globe is facing the pandemic and it has devastated the life of society. On the grounds of diverse medical facilities and variants of COVID 19 cases the proposed solution provides in depth analysis of incidents generated information, which is information of COVID 19 patients which maximizes benefits to health organizations and also maximizes value relationship with its medical stakeholders. This gives rise a thirst for carrying out the study on the available data analysis of COVID 19 data and use of some machine learning techniques to support decision making by identifying patterns using fuzzy classifications, dominance analysis by using Principle component analysis and to provide machine learning based predictions on available data for COVID 19 cases specifically for severity level and level of recurrence with respect to different input parameters(Age, time of diagnosis, pre-existing conditions, and symptoms). So, the study focuses to identify relationships, patterns and dominance which will support COVID 19 medical research field.

**Keywords:** COVID 19, Dominance Analysis, Fuzzy Logic, Machine Learning.

## 1. INTRODUCTION

Respiratory viral infections distress specifically lungs, nose, and throat. Such infection normally spread by breathe in droplets having virus particles. Spanish flu was viral disease of 1918 which infected 500 million people that is around one third population of planet. And now the novel virus Coronavirus disease (COVID-19) also infected to large population of the world. Most people who got infected with COVID-19 experienced mild to moderate symptoms and recover without special treatment, but so many patients lost their lives because of not exact, timely and proper treatments or existing medical problems and other threatening parameters. The COVID-19 has banged the limitations of decision making in Respiratory viral infections treatments and decision making at different level of severity of such diseases.

Technical solution can address different problems in manual decision making by analyzing voluminous data on few clicks can be combined with expertise skills and knowledge of medical professionals for treatment and decision making in such respiratory viral infections.

How it spreads

The virus that causes COVID-19 is mainly transmitted through droplets generated when an infected person coughs, sneezes, or exhales. These droplets are too heavy to hang in the air, and quickly fall on floors or surfaces.

Address for correspondence: Manisha Shinde-Pawar,  
Institute of Management and Rural Development Administration, Sangli  
Email: [mjs.imrda@gmail.com](mailto:mjs.imrda@gmail.com)

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: [pnrjournal@gmail.com](mailto:pnrjournal@gmail.com)

How to cite this article: Manisha Shinde-Pawar, Rajendra Pujari, Ayesha Mujawar, Alok Shah, Deepali Gala, Bhaskar Patil, Application of Machine Learning for COVID-19 Data Analysis, J PHARM NEGATIVE RESULTS 2022;13:334-338.

### Access this article online

#### Quick Response Code:



Website:  
[www.pnrjournal.com](http://www.pnrjournal.com)

DOI:  
10.47750/pnr.2022.13.03.053

You can be infected by breathing in the virus if you are within close proximity of someone who has COVID-19, or by touching a contaminated surface and then your eyes, nose or mouth.

#### Symptoms

COVID-19 affects different people in different ways. Most infected people will develop mild to moderate illness and recover without hospitalization.

Most common symptoms:

1. fever.
2. dry cough.
3. tiredness.

Less common symptoms:

1. aches and pains.
2. sore throat.
3. diarrhea.
4. conjunctivitis.
5. headache.
6. loss of taste or smell.
7. a rash on skin, or discoloration of fingers or toes.

Serious symptoms:

1. difficulty breathing or shortness of breath.
2. chest pain or pressure.
3. loss of speech or movement.

Till today there is no certain medicine or treatment is available, hence it is essential to create awareness among the people about covid-19 symptoms, its treatment and how it impacts to various body organs after the treatment and what is the probability of recurrence of disease.

#### Digital Support / Solution to get the track of treatment

An existing research study helped to learn techniques and framework for fuzzy classification, for dominance analysis and implementation and machine learning techniques applied in different areas.

#### Fuzzy Logic

Fuzzy Logic (FL) is a method of reasoning that resembles human reasoning. This approach is similar to how humans perform decision making. And it involves all intermediate possibilities between YES and NO. The Fuzzy logic works on the levels of possibilities of input to achieve a definite output.

One legacy artificial and machine learning technology is fuzzy logic. Traditional and classical logic typically categorize information into binary patterns such as: yes/no, true/false, or day/night. Fuzzy logic instead focuses on characterizing the space between these black-or-white scenarios.

#### Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

#### Dominance Analysis

Whenever multiple regression is used to test and compare theoretically motivated models, it is of interest to determine the relative importance of the predictors. Specifically, researchers seek to rank order and scale variables in terms of their importance and to express global statistics of the model as a function of these measures. It reviews the many meanings of importance of predictors in multiple regression, highlights their weaknesses, and proposes a new method for comparing variables: dominance analysis. Dominance is a qualitative relation defined in a pairwise fashion: One variable is said to dominate another if it is more useful than its competitor in all subset regressions.

#### Rationale and Significance of the Study

The proposed study will be significant for fuzzy based algorithm for impact classification and pattern detection to support COVID 19 treatment decision making.

- This study primarily determines various possible input parameters influencing COVID 19 treatment.
- The classification and patterns will help medical researchers and experts to identify tracks for treatment, priorities and segregation of patients. The study classifies impact into Serious, Average and deferred cases so accordingly research can focus to class specific symptoms and accordingly treatment can be planned.
- Dominance analysis will help to identify dominance of input variables on Health condition of a Patient with reduced dimension to get real insights of data spread.
- As incident based data captured will be transformed to contextual information will assist to design strategic policies with additional importance in all fields of COVID-19 research and impact.

## 2. A Brief about the Present Research Status in Domain under Consideration (Literature Review)

After studying current ML applications for range of clinical outcomes, prediction, diagnosis and reporting performance, authors Suggested need of improvement in ML applications especially in diverse health settings, primary care and routine clinical care environments [9].

According to authors[5], world’s critical situation in the current pandemic demands for more comprehensive and optimized ML approaches for medical treatment research, patient care, resource management and there is need of Emergency ML to support in the shocking consequences.

Authors suggested that, both population-wide and custom-made approaches should be developed as infectious diseases have diverse infectivity, incubation time, transmission mode, and will lead to different signs dependent on the host [10].

After Applying ML to number of clinical reports [6] [8], researchers could classify these reports into four classes and showed improved results of performance measures of interestingness as accuracy, precision, recall and f-score and suggested improvement for better version of ML models with focus on feature engineering and deep learning for future work.

The recent publications about applications of ML approaches to COVID 19 suggested that advancement in Models can prevail most effective and accurate in managing and treating information, medical cost, and other parameters. Application of ML as intelligent systems can be advanced to mark independent decisions in pandemic to support medical practitioners [7].

#### Research Gap

After going through the preliminary research investigation, it seems as problem is very recent, very less work is addressed so more reliable and innovative machine learning-models can be developed. The pandemic triggered immediate need of possible solutions to support medical resources and stakeholders. It is challenging uncertainty for the entire world to come out with the optimised immediate solutions which is attracting academicians and researchers to address the problem.

### 3. The objectives are to study

- i. To identify pattern and association of covid-19 symptoms and its impact on body organs.
- ii. To study dominance in symptoms of covid-19
- iii. To study and identify relationship with respect to different parameter with level of severity and probability of recurrence.

### 4. Tentative Research Methodology

The researcher plan to follow Design and Creation research Strategy. The strategy focuses on design of fuzzy and machine learning algorithm for COVID-19 data analysis, classification and relative study.

#### 4.1 Data Collection:

Primary Data: Bharati Hospital, Sangli

Secondary Data: News, articles, Research articles, from Media like websites, New Papers, TV etc.

Thus, researcher would like to study and experiment by implementing approach as given here to get improved design of fuzzy algorithm for sentiment analysis of social

media data.

#### 4.2 Scope of the Study

- Irrespective of varied usefulness of COVID-19 patient’s data, researcher has planned to application of machine learning and related blended technological solutions to COVID-19 data for selected input parameters only.
- Major Technological issues identified will be addressed
- For research scope of area will be restricted to Sangli city

#### 4.3 Software Used:

Software used for programming and implementation in windows-7 environment are as shown in Table No. 1

**Table 1** Software Used

Sr. No.	Software Used	Version
1	R Tools Framework	3.5.2
2	R Studio	1.1.463
3	Google Forms	Version 103.0.5060.66 (Official Build) (32-bit)
4	Microsoft Excel	Office 2007

Source: Compiled by Researcher

#### 4.4 Research Motivation

After studying all this situation of COVID-19, one can judge the future impact of such viral infections. Main challenge of Covid-19 is that the symptoms are changing from patient to patient as days passes so it’s very difficult to set line of treatments. And also new trends are found in European countries.

So, there should be an ICT based model that will take patient data as an Input and after processing that data it should generate some results which could be useful for the doctors for deciding future line of treatment for patient based on the patients’ health condition or severity of Infection.

So, The Researcher has identified some technical challenges for analysis of data through preliminary investigation and observation. To explore unfamiliar, exploration of new patterns and to scrutinize exceptions, incidents, cases by using strategic, structured and organized value mining to reach at sound decision is motivating potency for research in COVID 19.

To deal with vague data, fuzzy logic has played vital role in different domains. So, by considering the need of COVID research and treatment different analysis tools and algorithms for COVID can be focused to design in an action on existing data.

#### 4.5 Statement of Problem

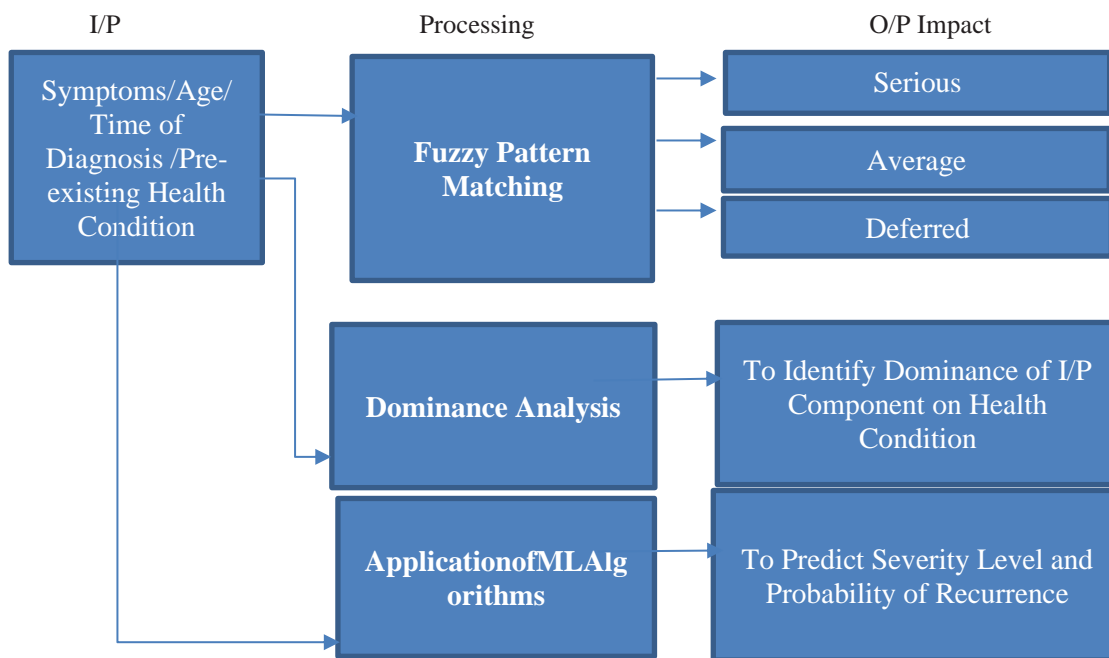
The researcher proposed to study tendency of data to achieve refined meaningful classes, patterns and groups by applying fuzzy based classification of impact, machine learning for

prediction and PCA for dominance grouping.

The researcher aims to address the limitations of data analysis by proposing study titled,

“Application of Machine Learning for COVID-19 Data Analysis.”

#### 4.6 Proposed Research Model



**Figure No. 1** Proposed ML Framework for COVID-19 Data Analysis

### 5. Research Project Impact

(Techno-Commercial status / Outcome / IPR / Social benefit /other) -Expected outcome

The research outcome for proposed work can be as follows

#### 5.1 Decision Making Model -

The Fuzzy classification model can help to segregate and plan the set of clinical treatments, experts, necessities and other resources. This can be identified as Techno-commercial advantage in medical field.

#### 5.2 Solution to Post COVID consequences –

Post COVID problems rising with its variants need to addressed at its initial level to avoid its severe impact or losses. Proper data grouping and dominance analysis also can be useful to understand the impact and level of exposure of different variants. This can help to plan social strategies and government policies for routine post COVID care, treatment and precautions.

#### 5.3 Implementation at Micro level units for day to day or routine healthcare setup design.-

Diverse health settings in large population countries like India and China is also challenge to provide medical treatments and facilities, create awareness in different kind of society, so micro level units and hospitals, clinics or

individual practitioners if would implement these automated solution at their service points or clinics, this will reduce the load on specialty hospitals and government hospitals.

#### 5.4 Hospital Management-

As the proposed work focuses clinical decision making support it will provide management solutions and techno-medical- management strategic models can be designed in future work for Hospital Management.

#### 5.5 Consideration for IPR –

The Proposed model is divided into 3 technical parts, Fuzzy Classification, Dominance Analysis and Evolutionary Model for prediction. Novel Part of Model can be considered for IPR registration in future implementation.

### 6. Conclusion:

In the research work, the researcher has planned and formed experimental strategy model for COVID-19 Data Normalization using text mining symptoms will be preprocessed and further in combination with other input parameters like Age, oxygen level, RT-PCR Test Result inferred for fuzzy classification for relative pattern and Impact of disease certainty analysis; graphical relative analysis, Coefficient of relationship and Principal Component Analysis will be applied to get insights of dimensions and to capture dominant group behavior, and Evolutionary model

and Fuzzy Clustering mean analysis for prediction. This application of machine learning approach work to COVID-19 data specifically is analysis in form relevance analysis, classification, prediction and grouping will support medical experts and doctors for timely decision making in treatment and care for viral respiratory diseases.

## 7. Acknowledgement

The Research work (Principal Author Dr. Manisha Shinde-Pawar) was financially supported by Bharati Vidyapeeth (Deemed to be University), Pune, Institute of Management and Rural Development Administration, Sangli.

## REFERENCES

[1]. Aljameel Sunayh, Khan Irfan, Aslam Nida, Aljabri Malak, Alsulmi , “Machine Learning Based Model to predict the disease Severity and Outcome in COVID 19 Patients,”Scientific Programming, Vol 2021

[2] Assaf, D., Gutman, Y., Neuman, Y. et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. *Intern Emerg Med* 15, 1435–1443 (2020). <https://doi.org/10.1007/s11739-020-02475-0>

[3]. Aydin N., Yurdakul G. Assessing countries’ performances against COVID-19 via WSIDEA and machine learning algorithms. *Appl. Soft Comput.* 2020; 97:1–18. Doi: 10.1016/j.asoc.2020.10679. Elsevier.

[4]Cascella, M., Rajnik, M., Aleem, A., Dulebohn, S. C., & Di Napoli, R. (2022). Features, evaluation, and treatment of coronavirus (COVID-19). *Statpearls* [internet].

[5]Ayoubkhani D, Khunti K, Nafilyan V, Maddox T, Humberstone B, Diamond I et al. Post-covid syndrome in individuals admitted to hospital with covid-19: retrospective cohort study *BMJ* 2021; 372 :n693 doi:10.1136/bmj.n693

[6] Debnath, S., Barnaby, D.P., Coppa, K. et al. “Machine learning to assist clinical decision-making during the COVID-19 pandemic”. *Bioelectron Med* 6, 14 (2020). <https://doi.org/10.1186/s42234-020-00050-8>

[7]. Khanday, A.M.U.D., Rabani, S.T., Khan, Q.R. et al. (2020) “Machine learning based approaches for detecting COVID-19 using clinical text data”. *Int. j. inf. tecnol.* 12, 731–739 (2020). <https://doi.org/10.1007/s41870-020-00495-9>

[8]. Mohammad Sadeq Mottaqi, Fatemeh Mohammadipanah, Hedieh Sajedi (2021), “Contribution of machine learning approaches in response to SARS-CoV-2 infection, Informatics in Medicine Unlocked”, Volume 23, 2021, 100526, ISSN 2352-9148, <https://doi.org/10.1016/j.imu.2021.100526>. (<https://www.sciencedirect.com/science/article/pii/S2352914821000162>).

[9]. Muhammad, L.J., Algehyne, E.A., Usman, S.S. et al. (2021) “Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset”. *SN COMPUT. SCI.* 2, 11 (2021). <https://doi.org/10.1007/s42979-020-00394-7>

[10]. N. Peiffer-Smadja, T.M. Rawson, R. Ahmad, A. Buchard, P. Georgiou, F.-X. Lescure, G. Birgand, A.H. Holmes (2020), “Machine learning for clinical decision support in infectious diseases: a narrative review of current applications, *Clinical Microbiology and Infection*”, Volume 26, Issue 5, 2020, Pages 584-595, ISSN 1198-743X, <https://doi.org/10.1016/j.cmi.2019.09.009>. (<https://www.sciencedirect.com/science/article/pii/S1198743X1930494X>)

[11]. Said Agrebi, AnisLarbi (2020), “Use of Artificial Intelligence in infectious diseases.”, *Artificial Intelligence in Precision Health*, Academic Press, 2020, Pages 415-438, ISBN 9780128171332, <https://doi.org/10.1016/B978-0-12-817133-2.00018-5>. (<https://www.sciencedirect.com/science/article/pii/B9780128171332000185>) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7153335/>

[12]. <https://www.datarevenue.com/en-blog/machine-learning-covid-19>

[13]. Osaka University. "Sorting out viruses with machine learning." *ScienceDaily*, 11 November

2020. <https://www.sciencedaily.com/releases/2020/11/20111102641.htm>

[14] <https://oig.hhs.gov/oei/reports/oei-06-20-00300.pdf>

[15]. <https://www.commonwealthfund.org/publications/2020/apr/impact-covid-19-outpatient-visits>