COVID-19 GROUPING INSIGHTS – PRINCIPAL COMPONENT ANALYSIS

[1*]Manisha Shinde-Pawar, [2]Jagadish Patil, [3]Alok Shah, [4]Pallavi Jamsandekar and [5]Dexter Woodward

[1,2,5]Assistant Professor, [4]Professor and I/C Director, Bharati Vidyapeeth (Deemed to be University), Institute of Management and Rural Development Administration, Sangli

[3]Assistant Professor, Bharati Vidyapeeth (Deemed to be University), Department of Management Studies, Navi Mumbai

[1]*mjs.imrda@gmail.com

[1]*ORCID- 000-0002-8276-9782,

[1]*Web of Science Researcher ID: GSD-3134-2022

ABSTRACT

Pattern and Group behavior of COVID-19 patient's data is very much important aspect of data analysis for medical stakeholders to frame decision strategies and to design routine set-ups. PCA is old mathematical machine learning grouping analysis technique but popular in recently for data analysis in grouping insights, so the researcher has implemented principal component analysis to study dominance in data with varied component grouping with proportional variation and also used graphical visualizations. The strength of PCA implementation is to maintain real valued data with dimension reduction but without loss of key information. The experiment result of COVID-19 datais showing component 1 contributing to highly positive testResult that is having High impact on patients. PCA shows relative dimensions analysis. PCA dimensions plotted shows independence in dimension with 90^0 angle dimensions in data spread. The results identified for COVID-19 data glimpse with dominant component formation.

Keywords: C O V I D - 1 9 , Grouping Insights, Pattern Detection, PCA: Principal Component Analysis

1. INTRODUCTION

Principal Component Analysis (PCA) implemented to five variables Age, CT Score, O2 level, Severity Level and RT-PCR Test Result of pre-processed COVID-19 dataset of 200 data objects for dimension reduction and dominant data group construction. PCA Summary and PCA loading values are obtained and major components are plotted using histogram for COVID-19 dataset. Further scree plot also depicts variance reliability and relative analysis and biplot for dimension relationship analysis are plotted for the obtained principal components.

2. LITERATURE REVIEW

- **Machine Learning For COVID-19 Treatment**

After carrying systematic review of 15 different papers authors [R-4] found that, the **use of efficient machine learning based clinical decision support system for arboviral diseases can improve the quality** of the entire clinical process, thus **increasing the accuracy of the diagnosis and the associated treatment.** And**current research for clinical data is limited to the three most common infections—Dengue, Chikungunya and Zika.**

According to authors [R-7], the application of AI algorithms on covid-19 research is at its infancy and there is still much room for improvement and new areas that AI can be used in tackling the problem. It is important to know how the problems caused by the disease would look like in future to be able to plan strategies from now as they may be easier to cope with at their infancy. According to authors[R-15], world's critical situation in the current pandemic demands for more comprehensive andoptimized ML approaches for medical treatment research, patient care, resourcemanagement and there is need of **Emergency ML** to support in the shocking consequences. Application of ML as intelligent systems can be advanced to mark independent decisions in pandemic to support medical practitioners [R-16].

R-12] After studying current ML applications for range of clinical outcomes, prediction, diagnosis and reporting performance, authors Suggested need ofimprovement in ML applications especially in diverse health settings, primary care and routine clinical care environments [R-13]. This study [R-3] aimed to assess the utility of machine-learning algorithms for predicting outcome of patients with non- critical COVID-19 based on clinical parameters on admission. We found clear difference between patients who developed later critical vs. non-critical COVID-19, mainly in **vital signs (respiratory rate and room-air oxygen saturation) and inflammation markers (blood WBC, neutrophil counts and CRP)** and in the APACHE II score that combines these makers. Authors assessed clinical, hematological and biochemical parameters at admission. [R-2] The prognosis of COVID-19 is largely dependent on various factors that include **the patient's age, the severity of illness at presentation, pre-existing conditions, how quickly treatment** can be implemented, and response to treatment.

Severity Level: Most of the datasets used to have two labels for the possible output classes (i.e., infected or not infected) [R-39]. Authors suggested need to have a dataset that is annotated to show the severity level of the illness to monitor patient healing progress. **Effect on different body organs**: Most of the studies rely on investigating lung or chest images to predict the infection with COVID- 19. Other organs affected by the virus need to be investigated further as this may ease the virus detection and reduce the risk of being infected and not knowing that. Monitoring **Patient History, Vaccine status, Lifespan of virus, effect on body** these findings can help to identify and prevent patterns.

- **PCA for COVID 19 treatment**

In PCA Analysis of data populated [R-40], Authors identified largest cluster having around 53% data components of COVID patients with features insights as oldest age, less immunity, and less albumin levels, whereas very less only 9% data captured in group of third cluster represented data of patients who were having very young age and high immunity power features. Authors[R-9] compared different countries and spread rate in high risk countries also suggested classify COVID-19 data using factoranalysis for other countries.

- **Analysis using R**

The impetration of COVID-19 data research using R mentioned limitation of large corpus volume, which can be handled by using machine learning methods available in R[R-21].

3. Objectives

The main purpose of current study by using grouping insights is to apply fuzzy based pattern of COVID-19 media data using R tool that would help the expert, professionals, governments as well as students in the process of medical treatment and care.

The objectives of present research were to collect the COVID-19 patients' treatment and pre and Post COVID data and parameters required for the study, accordingly to collect from Bharati Hospital, Sangli and to apply fuzzy rule-based system to COVID-19 Data for right classification, prediction and grouping and supremacy analysis. For the said research work the researcher has set following objectives:

- To identify pattern and association of covid-19 symptoms.
- To study dominance in constraint parameters of covid-19 data.

4. GROUPING INSIGHTS

Principal Component Analysis COVID19 Patient's Data Analysis

4.1 PCA for COVID19 Patient's Data Analysis:

COVID-19 disease patient's data is huge dataset with multiple possible dimensions for analysis. Though PCA is old Mathematical model, in modern era of data analysis PCA has become active and quite precise extensively being used analysis technique for dimension reductions. By keeping variably

ISSN 2515-8260 Volume 09, Issue 07, 2022

in the data under analysis, PCA preserves content of statistical information with reduced or transformed data without loss of information content. G. Vinodhini, R. M. Chandrashekaran (2014) mentioned PCA as active a feature reduction method in combination with ANN.

4.2 PCA for COVID-19 Patient's Pre-Processed Data

For COVID-19 dataset five dimensions are scaled to obtain principal components forming five components with statically information as proportional variance 41%, Standard deviation for Major Component 1 is 1.44 which is maximum value related to other four components analysed. So the Component 1 shows the maximum variation seized.

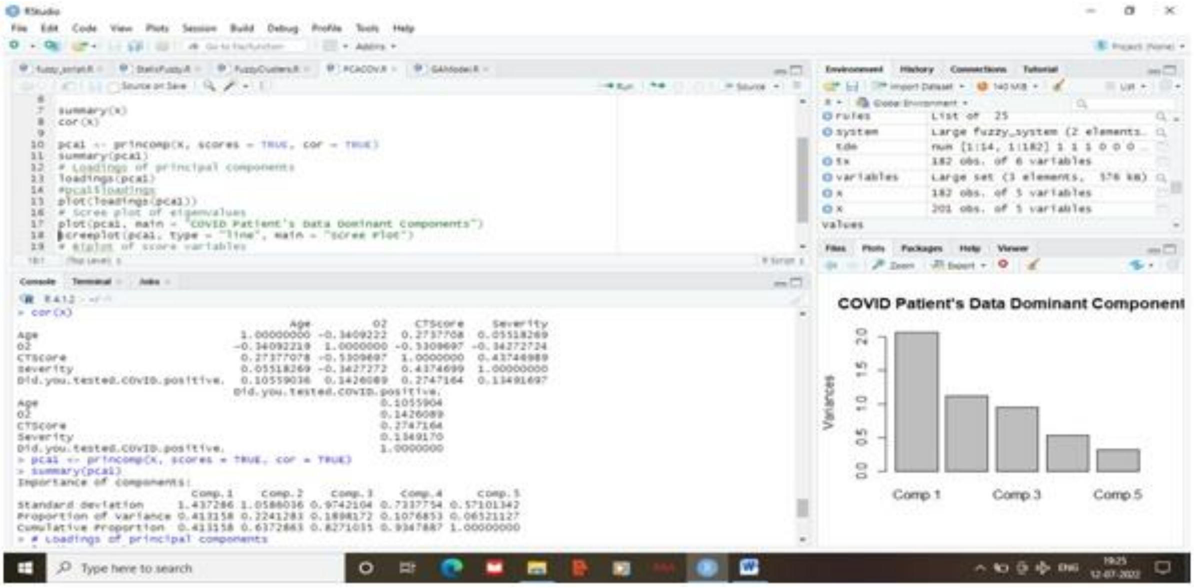

Figure 1: Component Analysis

Importance of components:	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	1.44	1.06	0.97	0.73	0.57
Proportion of Variance	**0.41**	**0.22**	**0.19**	**0.11**	**0.07**
Cumulative Proportion	0.41	0.63	0.82	0.93	1.00

Table 1

Component Variable	Component 1	Component 2	Component 3	Component 4	Component 5
Age	0.363	0.140	0.785	0.461	0.138
O2 Level	-0.533	-0.448	---	0.309	0.647
CT Score	0.590	-0.138	---	-0.486	0.626
Severity	0.454	-0.127	-0.577	0.666	---
COVID Positive	0.174	-0.863	0.209	-0.106	-0.413

PCA Loading Values of COVID-19 Dataset
Source: Compiled by Researcher

As mentioned in table no. 1, the PCA loading value close to 1 that is positive associationof dimension represents the majority of positive influence or values close to -1 are measured as maximum influenced negative association of dimension, so as component 1 with Age, CTScore, and Severity level shows values 0.363, 0.590, and 0.454 is closer to 1 shows large positive association with COVID RT-PCR TestResult variable, at the sametime instance Component 1 has O2 Level variable value -0.533 is closer to -1 which shows large negative or opposite association with COVID RT-PCR TestResult variable. That is Component 1 revokes large positive with COVID RT-PCR TestResult dominance in COVID-19 patient's records about infection and treatment and large negative association with Oxygen level that reveals that if Oxygen level drops possibility of positive TestResult increases. The component1 is major component formed showing dominance in Positive Test Result for increased Age, CTScore and Severity and for decreased O2 level, so based on

ISSN 2515-8260 Volume 09, Issue 07, 2022

COVID-19 dataset from sample collected, PCA results clearly shows the strong Positive RT-PCR Test infection results for increased values of Age, CTScore and severity and decreased values of O2 Level.

In Component 2 O2 level, CT Score, Severity value is - 0.45, -0.14 and -13 which is adjacent to -1 and positive COVID Test variable is -0.87 again close to -1 showsnegative association with COVID Test variable that represents less score of severity and CT Score though drops oxygen then also chances of infection are less. In Component 3, Age variable with loading value 0.79 and High positive association with severity shows almost doubled Positive RT-PCR Test Result association which yet again supports to Positive RT-PCR Test Result dominance. Component 4 also represents average Age, O2 Level and Severity and negative association CTScore represents less chances of getting Positive RT-PCR Test Result. Component 5 is having smallest preference due to low variation.

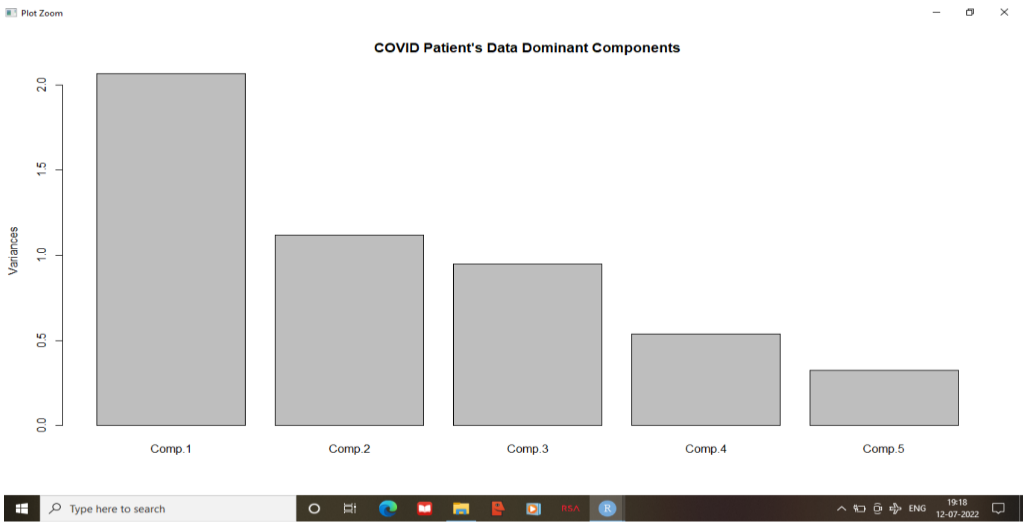

Figure 2: Histogram of Components-Variance.

Extreme dissimilarity in component detects key dominant components for analysis in case of huge voluminous data. The Figure 6.2 of Histogram plotted depicts Component1 with 41% variance, which reveals that majority of COVID-19 patient's data, could be summarized by means of Component1 as single major representative component. Component2 represents 22% of the total variance; Component3 obtained with 19% of the variance. Component4 and Component5 creates 11%, 7% variance respectively,therefore, Component1, Component2 and Component3 can express real glimpse of data in relation with other components, one can understand the major aspect, in all data, as the Component1, Component2 and Component3 explains 82% of the variance.

Without loss of valued data reduction in numbers of dimension focuses on dominant COVID-19 data and gives view to key influencing qualities of dataset. As depicted in Figure 3, the Scree plot shows relative variance proportion and relationship of variance value consistency of component showing variations in all components 1 to component5.

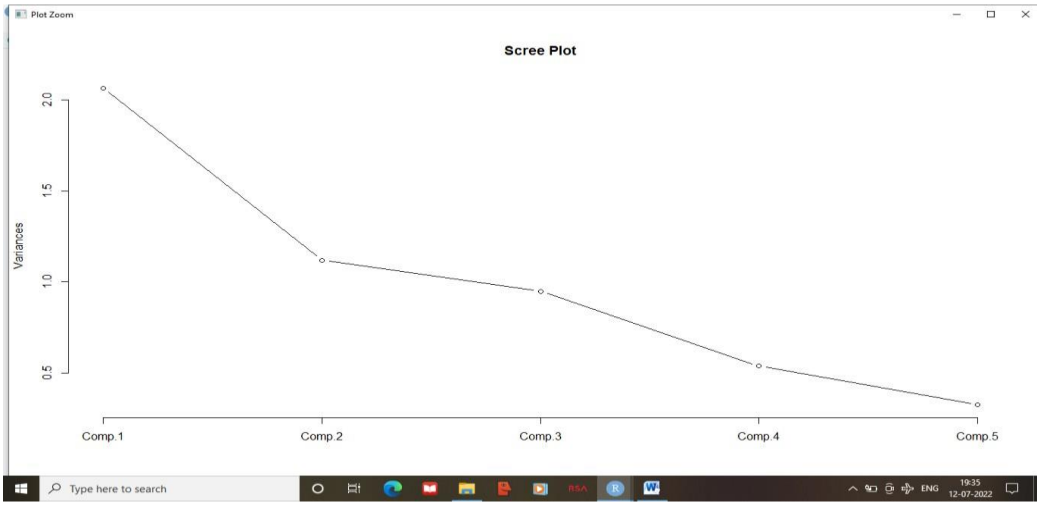

Figure 3: Scree Plot of Component Variances

The first component representation in Scree plot with maximum variation in relation to other four components can be identified as major or prime component constructed. Succeeding representative components are positioned screening the influence of quantity of variation each one is demonstrating. Therefore, the First Key Component 1 groups the major or principal part of distinction, and then the Component 2 is next large component of second largest variation, and so on sequentially. The Scree Plot depicts contribution of all the components with respect of variation amount estimated. Largest dissimilarity and the relative principal component signify to supreme distinctive qualities or behaviour of COVID 19 dataset. So skipping or missing information from the identified major components may be risky and can lose significant information.

First three key components are dominant components which captures majority of the valuable information centring on dominant data of Age, CTScore, Severity level, oxygen level and RT-PCR COVID Test Result value dimensions. As key Component Principal Component 1 provides exploration to most influencing features that are influencing to Positive or High values Age, O2 level, RT-PCR Test Result with maximum Low level O2 Level concentration. Severity and CT Score are at almost same level so one of these features can support the judgement of Test Result. Dimensions relationship in scree plot obtained are showing dominance in POSITIVE RT-PCR Test Result in dataset.

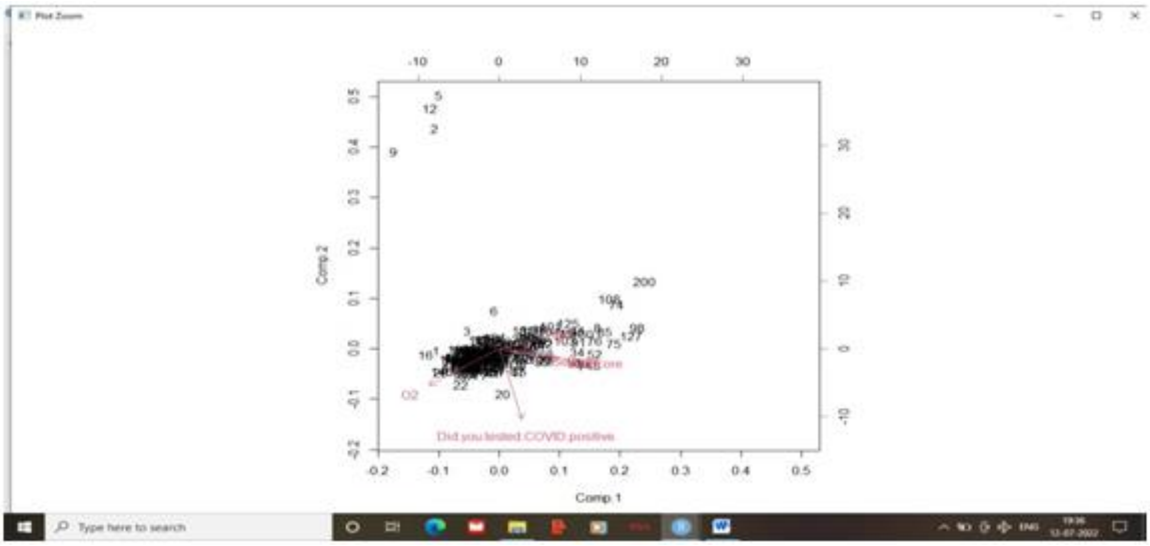

Figure 4: Biplot of Score Variables

Biplot denotes multivariate information background in which the matrix with top-right axes reflects loading values for the principal components formed and other axes shows score values for the principal components. The biplot dimension angles represents Age, O2 Level, CT Score or Severity Level and RT-PCR Test Result are independent dimensions required for COVID-19 data analysis.

A PCA Biplot in Figure 4 shows component PCA score of dimensions in black numbers and vector signifies loadings values. Biplot shows how much variation each principal component captures from the data spread. As three Principal Components are enough to gain essence of the data, the scree plot is a sharp arc that curves swiftly and levels out.

The figure 4 shows, all five variables but four Age, Severity and TestResult are almost in 90^o angle with each other, which indicates there is no dependence between loading values of O2 level-COVID Test and Age- COVID Test. And almost all three of these arescreening their inevitability in unfolding essence of data; hence O2 level and Age are highly important input parameters which are valid, CT Score and Severity level are at symmetrical level so either this along with first two O2 level and Age are sufficient parameters for information retrieval in COVID19 RT PCR Test Result. PCA viewing variation means for different qualities/features/characteristics which are important to analyze different COVID19 RT PCR Test results can be signified by means of these constraints.

5. OBJECTIVE FULFILLMENT

To study dominance in constraint parameters of covid-19 data.

Dominance Detection is applied and explained in form of principal components representing real glimpse of entire dataset. Fuzzy c-means also identifies degrees of central data membership for pattern matching and identifying major cluster formed of similar behavior.

6. FUTURE SCOPES FOR RESEARCH

In future research, deep learning can be applied to COVID-19 data set. Researcher would like to suggest developing AI based model for unstructured data also. As the pandemic and related data is very recent, respiratory disease and solutions are attractions of recent research, so can be applied with different expert and recommendation systems. The emergent investigation and ultimatum in research scholar positions for COVID is also screening the interest of scientific community. Investigator would like to propose enhancement in the work in collaboration with medical research fellows to focus on innovative methods for computerized medical support systems. This work can be further enhanced fuzzification of images and video identifying as new parameters and by simplifying it to standard form and allocating weights for more exact set of perditions. One can also apply image processing to HRCT scan image reports to accept it as major, vital input constraint.

7. CONCLUSION

In Dominance analysis using Principal Component Analysis statistical technique, five different components are produced for sample COVID-19 data. As PCA shown in section 6 PCA1 showed maximum data points representing strong Component 1 in result give an idea about dominance explaining maximum 41% proportional variance. Comp2 is next largest variance 22% variance and Comp3 third largest component with 19% variance. Component1 shows dominance of low oxygen level, high CT Score and severity associated with positive RT-PCR testResult whereas Component3 High Age Value with less or no severity also showed strong data dominance associated to strong positive RT-PCR testResult for COVID-19 pre-processed data records, so entire data with maximum variation associated to positive and strongly associated to positive RT- PCR testResult value. These results reflects that high Age value patient's not having highCT Score also resulted in either death or complications in treatment as compared to low Age value patients. That is patients with old age faced major risks and complications of COVID-19 infection, once infected. And Age and Impact can be analysed as proportional values. That is associated with each other. If Age increases, impact increases and vice versa. O2level and Impact are inversely proportional. That is negatively associated with each other. If O2level increases, impact decreases and vice versa.

In the research work, the investigator has investigated and formed strategy for COVID-19 Data Normalization, fuzzy classification for pattern and certainty analysis used;graphical relative analysis, Coefficient of relationship and Dominance analysis is used to recognize important input constraints for different approaches to balance variation as well as dominance in data to improve real crux of data. Different grouping techniques are used to study grouping insight which are useful for dominance detection; this application of machine learning approach work to COVID-19 data specifically is analysis in form relevance analysis, classification, prediction and grouping to support medical stakeholder'sdecision making in treatment and care for viral diseases.

8. ACKNOWLEDGEMENT

The Research work (Principal Author Dr. Manisha Shinde-Pawar) was financially supported under minor research project scheme 2021-22 by Bharati Vidyapeeth (Deemed to be University), Institute of Management and Rural Development Administration, Sangli.

9. BIBLIOGRAPHY

9.1 BOOKS

1. George Klir and Bo Yuan (2009), "Fuzzy sets and Fuzzy Logic, Theory and Applications", Book ISBN-978-81-203-1136-7 PHI Learning Pvt Ltd Indian Edition.
2. HrishikeshaMohanty, PrachetBhuyan, Deepak Chenthati(2015), "Big Data :A Primer" ISSN 2197-6503 ISBN-978-81-322-2493-8, by Springer Pvt Ltd. , Copyright-Springer India. 5. Steven Finlay(2014), "Predictive Analytics, Data Mining and Big Data : Myths, Misconceptions and Methods",ISBN-978-1- 137-37927-6, Palgrave MacMillan

9.2 THESIS

1. Christian Infante (2017), "Development of Machine Learning Algorithms for Screening of Pulmonary Disease", MASSACHUSETTS INSTITUTE OF TECHNOLOGY.
2. Filip Velickovski(2016), "Clinical Decision Support for screening, diagnosis and assessment of respiratory diseases: Chronic Obstructive Pulmonary Disease as a use case", University of Girona
3. Manisha Shinde (2020)," Effectiveness Fuzzy based Sentiment Analysis in measuring impact of Social Media with special reference to Facebook and Twitter", Bharati Vidyapeeth (Deemed to be University), Pune

9.3 ARTICLE 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. Sof Comput. 2020; 97:1–18. Doi: 10. 1016/ j.asoc. 2020.10679. Elsevier.
4. da Silva Neto SR, Tabosa Oliveira T, Teixeira IV, Aguiar de Oliveira SB, Souza Sampaio V, Lynn T, et al. (2022) Machine learning and deep learning techniques to support clinical diagnosis of arboviral diseases: A systematic review. PLoS Negl Trop Dis 16(1): e0010061.
5. 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
6. Fujun Peng, Lei Tu, Yongshi Yang, Peng Hu, Runsheng Wang, Qinyong Hu, Feng Cao, Taijiao Jiang, Jinlyu Sun, Guogang Xu, Christopher Chang, "Management and Treatment of COVID-19: The Chinese Experience", Canadian Journal of Cardiology, Volume 36, Issue 6, 2020, Pages 915-930, ISSN 0828-282X
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. LiqiangGeng, Howard Hamilton (September 2006), "Interestingness Measures for Data Mining:

A Survey", ACM Computing Surveys, Vol. 38, No. 3, Article 9.

9. Mahmoudi , Mohmad and Heydari, Mohmmad Hpssein & Qasem, Sultan & Mosavi, Amir. ," Principal Component Analysis to study the relations between the spread rates of COVID 19 in High Risks Countries.", Alexandra Engineering Journal 2020

10. Mahdavi M, Choubdar H, Zabeh E, Rieder M, Safavi-Naeini S, Jobbagy Z, et al. (2021) A machine learning based exploration of COVID-19 mortality risk. PLoS ONE 16(7): e0252384. https://doi.org/10.1371/journal.pone.0252384

11. Marie Gousse, Pauline Penot, […] COCOREC study group." Clinical recurrences of COVID-19 symptoms after recovery: Viral relapse, reinfection or inflammatory rebound?" The Journal of Infection. Nov 2020, 81[5]

12. Mohammad-H. Tayarani N., Applications of artificial intelligence in battling against covid-19: A literature review, Chaos, Solitons & Fractals, Volume 142, 2021, 110338, ISSN 0960-0779, https:// doi.org/ 10.1016/ j.chaos.2020.110338.(https:// www. Science direct.com /science/ article/pii/S0960077920307335)

13. Mohammad Sadeq Mottaqi, Fatemeh Mohammadipanah, HediehSajedi (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

14. Mojjada, R. K., Yadav, A., Prabhu, A. V., & Natarajan, Y. (2020). Machine Learning Models for covid-19 future forecasting. Materials today. Proceedings, 10.1016/j.matpr.2020.10.962. Advance online publication. https://doi.org/10.1016/j.matpr.2020.10.962

15. MojdehJalali-Heravi, Osmar R. Zaïane(2010), "A Study on Interestingness Measures for Associative Classifiers", University of Alberta, Canada.

16. 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

17. N. Peiffer-Smadja, T.M. Rawson, R. Ahmad, A. Buchard, P. Georgiou, F.-X. Lescure, G. Birgand,

18. 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.

19. Norah Alballa, Isra Al-Turaiki, "Machine learning approaches in COVID 19 diagnosis, mortality, and severity risk prediction: A review", Informatics in Medicine Unlocked, Volume 24, 2021. ISSN 2352-9148

20. Osaka University. "Sorting out viruses with machine learning." ScienceDaily, 11 November 2020.https://www.sciencedaily.com/releases/2020/11/201111102641.htm

21. Osama R. Shahin, Hamoud H. Alshammari, Ahmed I. Taloba, Rasha M. Abd El-Aziz, Machine Learning Approach for Autonomous Detection and Classification of COVID-19 Virus, Computers and Electrical Engineering, 2022, 108055, ISSN 0045-7906, https:// doi.org/ 10.1016/ j.compeleceng.2022.108055.

22. Park, M.S., Jo, H., Lee, H. l. . , "Machine Learning-Based COVID-19 Patients Triage Algorithm Using Patient-Generated Health Data from Nationwide Multicenter Database", Infect Dis Ther (2022).

23. Patricia Carracedo, Rosa Puertas, Luisa Marti, "Research lines on the impact of the COVID-19 pandemic on business. A text mining analysis", Journal of Business Research, Volume 132, 2021, Pages 586-593, ISSN 0148-2963, https://doi.org/10.1016/j.jbusres.2020.11.043.

24. Peltan ID, Beesley SJ, Webb BJ, Lopansri BK, Sinclair W, Jacobs JR, et al. (2021), "Evaluation

of potential COVID-19 recurrence in patients with late repeat positive SARS-CoV-2 testing", PLoS ONE 16(5): e0251214.

25. Pijush Dutta, Shobhandeb Paul, Asok Kumar, Chapter 25 - Comparative analysis of various supervised machine learning techniques for diagnosis of COVID-19, Editor(s): Suman Lata Tripathi, Valentina E. Balas, S.K. Mohapatra, Kolla Bhanu Prakash, Janmenjoy Nayak, Electronic Devices, Circuits, and Systems for Biomedical Applications, Academic Press, 2021, Pages 521-540, ISBN 9780323851725, https://doi.org/10.1016/B978-0-323-85172-5.00020-4. (https:// www. Science direct.com/ science/article/pii/B9780323851725000204)
26. Prof. BhushanTrivedi(2017-2018), "Artificial Intelligence-I", Module 9, Module 36 by Prof. BhushanTrivedi, Swayam Courses 15th Aug 2017 to 15th Jan 2018.
27. Quiroz-Juárez MA, Torres-Gómez A, Hoyo-Ulloa I, León-Montiel RJ, U'Ren AB. Identification of high-risk COVID-19 patients using machine learning. PLoS One. 2021 Sep 20;16(9):e0257234. doi: 10.1371/journal.pone.0257234. PMID: 34543294; PMCID: PMC8452016.
28. S. Kanan and S. Bhaskaran(2009), "Association Rule Pruning based on Interestingness Measures with Clustering", IJCSI International Journal of Computer Science Issues, Vol. 6, No. 1.
29. 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.
30. Shaban, W. M., Rabie, A. H., Saleh, A. I., & Abo-Elsoud, M. A. (2021). Detecting COVID-19 patients based on fuzzy inference engine and Deep Neural Network. Applied soft computing, 99, 106906. https://doi.org/10.1016/j.asoc.2020.106906
31. Shinde-Pawar Manisha, Jagadish Patil, Alok Shah, Prasanna Rasal, "Development Of Fuzzy Inference System For Covid-19 Data Analysis", J Pharm Negative Results 2022;13: 693- 699
32. Toshiki Kuno, Yuki S, Shinpei K, Mai T, Masao I., Natalia N. E., "Prediction of in-hospital mortality with machine learning for COVID-19 patients treated with steroid and remdesivir", Journal of Medical Virology, Vol 94, Issue 3.
33. Wang Y., Hu M., Li Q., Zhang X.-P., Zhai G., Yao N. "Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with covid-19 in an accurate and unobtrusivemanner." 2020b. arXiv:2002.05534.
34. Warda M. Shaban, Asmaa H. Rabie,Ahmed I. Saleh, M.A. Abo-Elsoud[2021], "Detecting COVID-19 patients based on fuzzy inference engine and Deep Neural Network", Appl Soft Computing.
35. Weston C. Roda, Marie B. Varughese, Donglin Han, Michael Y. Li, Why is it difficult to accurately predict the COVID-19 epidemic?,Infectious Disease Modelling, Volume 5, 2020, Pages 271-281, ISSN 2468-0427,https://doi.org/10.1016/j.idm.2020.03.001. (https:// www. sciencedirect.com/ science/article/pii/S2468042720300075)
36. W.T. Li, J. Ma, N. Shende, G. Castaneda, J. Chakladar, J.C. Tsai, et al. Using machine learning of clinical data to diagnose covid-19 medRxiv (2020)
37. X. Wu, H. Hui, M. Niu, L. Li, L. Wang, B. He, et al. Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: a multicentre study, Eur J Radiol, 128 (2020), p. 109041
38. Xiangji Huang(March 2007), "Comparison of interestingness measures for Web usage mining: An empirical study", International Journal of Information Technology and Decision Making 06(01):15-41 ·
39. Y. Mohamadou, A. Halidou, P.T. Kapen ,[2020] A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of covid-19. Applied Intelligence (2020), pp. 1-13

40. Ye Wenjing, Lu Weiwei, Tang Yanping, Chen Guoxi, Li Xiaopan, Ji Chen, Hou Min, Zeng Guangwang, Lan Xing, Wang Yaling, Deng Xiaoqin, Cai Yuyang, Huang Hai, Yang Ling, “Identification of COVID-19 Clinical Phenotypes by Principal Component Analysis-Based Cluster Analysis“, Frontiers inMedicine, ISSN=2296-858X, Volume=7, Year=2020,URL=https:// www.frontiersin.org/articles/10.3389/fmed.2020.570614, DOI=10.3389/ fmed. 2020.570614
41. Z. Alyasseri, Al-Betar, M. A., Doush, I. A., Awadallah, M. A., Abasi, A. K., Makhadmeh, S. N., Alomari, O. A., Abdulkareem, K. H., Adam, A., Damasevicius, R., Mohammed, M. A., & Zitar, R. A. (2021). Review on COVID-19 diagnosis models based on machine learning and deep learning approaches. Expert systems, e12759. Advance online publication. https:// doi.org/ 10.1111/exsy.12759
42. Zoabi, Y., Deri-Rozov, S. & Shomron, N. “Machine learning-based prediction of COVID-19 diagnosis based on symptoms”. npj Digit. Med. 4, 3 (2021) https://www.datarevenue.com/en-blog/machine-learning-covid-19

WEBSITE REFERENCES

1. http://en.wikipedia.org/wiki/Prescriptive_analytics
2. http://en.wikipedia.org/wiki/Big_data
3. http://en.wikipedia.org/wiki/Artificial_intelligence
4. http://www.cognizant.ch/SiteDocuments/Cognizanti_Issue8_Big-Data_v2.pdf
5. http://www.sciencedirect.com/science/article/pii/S2090447914000550
6. http://www.ijettcs.org/Volume2Issue2/IJETTCS-2013-03-30-051.pdf
7. http://research.ijcaonline.org/volume99/number17/pxc3898243.pdf
8. http://ir.uiowa.edu/cgi/viewcontent.cgi?article=3090&context=etd
9. http://ac.els-cdn.com/S2090447914000550/1-s2.0-S2090447914000550- main.pdf? _tid= 32816eb8 -91de-11e5-8ac8- 00000aab0f 01&acdnat= 144828210 2_24a5ab8c 5f5c24c5b1 41ce cc5b1d 6c57
10. http://optima.jrc.it/wassa2015/ppts/wassa2015_submission1.pdf
11. http://www.ibm.com/developerworks/library/ba-sentiment-analysis-big-data/ba-sentiment-analysis-big-data-pdf.pdf
12. http://www.itu.dk/people/rao/pubs_accepted/2014_EDOC_BigSocialData_Fuzzysets.pdf
13. https://doc.rero.ch/record/30954/files/00002292.pdf
14. http://www.macs.hw.ac.uk/cs/techreps/docs/files/HW-MACS-TR-0033.pdf
15. https://researchportal.port.ac.uk/portal/files/4997381/ICACI_Paper_2.pdf
16. https://en.wikipedia.org/wiki/Data_pre-processing
17. https://towardsdatascience.com/the-ultimate-guide-to-data-cleaning-3969843991d4
18. https://en.wikipedia.org/wiki/R_(programming_language)#Programming_features
19. https://en.wikipedia.org/wiki/RStudio
20. https://www.r-bloggers.com/what-are-the-best-machine-learning-packages-in-r/
21. https://apacheignite.readme.io/docs/genetic-algorithms
22. https://en.wikipedia.org/wiki/Genetic_algorithm#Limitations
23. https://en.wikipedia.org/wiki/Machine_learning
24. www.bigdata-madesimple.com
25. www.swayam.gov.in , Artificial Intelligence module 9 by Prof.Bhushan(2018)

26. https://www.cdc.gov/coronavirus/2019-ncov/global-covid-19/compare-digital-tools.html