Early Diagnosis of Covid-19 From Routine Laboratory Tests (Machine Learning Model)

Youssef Shanan (*Faculty of Business Informatics*) / **Supervisors: Dr. Ayman Al Serafi & Dr. Sally Ibrahim / The German University in Cairo**

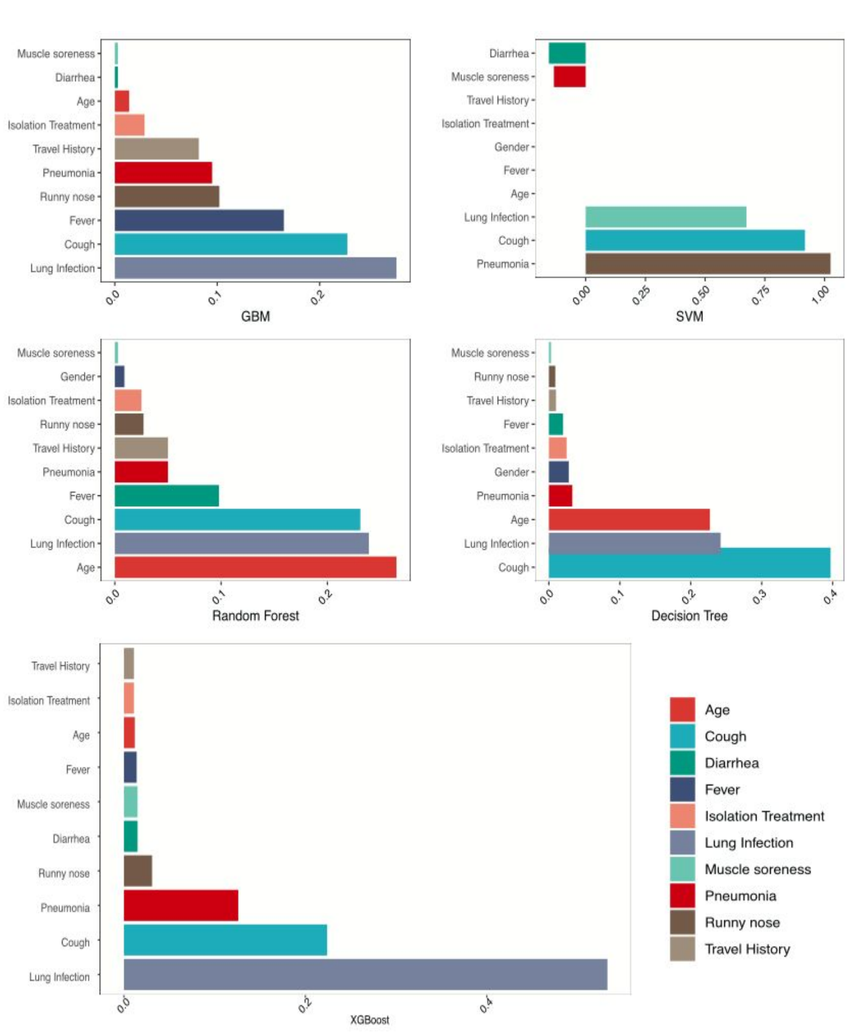
youssef.shanan@student.guc.edu.eg / ayman.alserafi@guc.edu.eg / sally.ibrahim@guc.edu.eg

Introduction

To begin with, the main objective of this study is to find a cheaper alternative to the RT-PCR test. The reason why another alternative was sought is that issues including sample infection and poor techniques for collecting, managing, transferring, and preserving samples, can significantly reduce the accuracy of the RT-PCR test (Brinati et al., 2020: 10). Consequently, this study aims to investigate the result of using machine learning models on the early diagnosis of COVID-19.

Literature Review

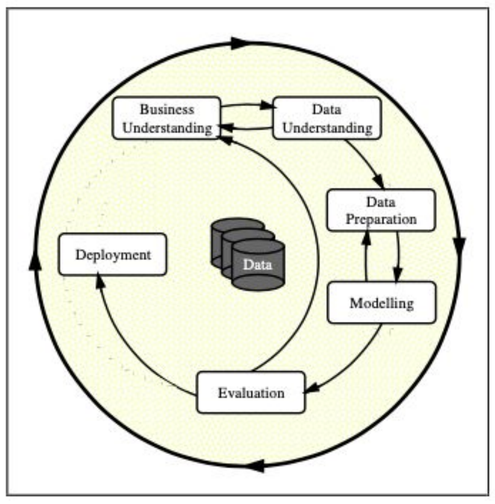
The discovery of tools to detect affected people as early as possible is critical to halting the spread of the disease (Ahamad et al., 2020: 8). Therefore, machine learning algorithms offer a potential solution that can be implemented quickly and cost effectively in a pandemic crisis (Ahamad et al., 2020: 8). Ahamad et al., (2020: 8) constructed and evaluated a number of machine learning models which revealed the main medical COVID-19 predictive characteristics: lung infection, cough, pneumonia, runny nose, travel history, fever, isolation, age, muscle soreness, diarrhea, and gender. The characteristics are written in a descending order.



[Characteristic importance for COVID-19 patients] [Ahamad et al., 2020: 8]

Methodology

The tool that was used in this study is the cross industry standard process for data mining which is known as the CRISP-DM. The CRISP-DM provides a structured approach for solving a problem by using data mining.



The dataset used in this study was extracted from a platform found on the internet called Zenodo.org. Furthermore, the platform was consulted by Dr. Federico Cabitza of University of Milano-Bicocca who is the creator of the dataset along with a group of other researchers. The dataset consisted of 279 records of patients admitted to the San Raffaele Hospital (Milan, Italy).

Results

A

At first, a stratified cross validation approach was used to train and assess the models indicated below. The loop analyses the efficiency of the model on 10 folds (Brinati et al., 2020: 10). Furthermore, the average of the scores of the 10 folds for every model was taken and added into the below table.

Model	Mean Cross-Validation F1-Score
Random Forest	0.807
Logistic Regression	0.785
K-Nearest Neighbors	0.736
Decision Tree	0.705

B

As demonstrated in the below table, four evaluation metrics were used to assess every model: accuracy, F1-score, precision, and recall. Furthermore, results concerning each model are added into the below table.

Model	Accuracy	F1-Score	Precision	Recall
Random Forest	0.798	0.862	0.828	0.898
Logistic Regression	0.750	0.824	0.817	0.898
K-Nearest Neighbor	0.702	0.786	0.793	0.780
Decision Tree	0.726	0.810	0.790	0.831

Conclusion

In conclusion, the study has showed how is it important in the current time to acquire artificial intelligence in the healthcare sector, as it is a rising technology that needs be to relied on in order to gain efficiency and cost savings. Furthermore, a research gap has emerged by Brinati et al., (2020:10) proposing that further research is required to explore other approaches for detecting COVID-19. In this study, a research gap was filled through proposing and constructing an alternative to the RT-PCR which is a machine learning model that is trained and tested with routine lab tests to be able to early detect COVID-19 in a cheaper and faster way.

References

1. Ahamad, M. M., Aktar, S., Rashed-Al-Mahfuz, M., Uddin, S., Liò, P., Xu, H., ... & Moni, M. A. (2020). A machine learning model to identify early stage symptoms of SARS-Cov-2 infected patients. Expert systems with applications, 160, 113661.

2. Brinati, D., Campagner, A., Ferrari, D., Locatelli, M., Banfi, G., & Cabitza, F. (2020). Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study. Journal of medical systems, 44(8), 1-12.

3. Chen, M., & Decary, M. (2020, January). Artificial intelligence in healthcare: An essential guide for health leaders. In Healthcare management forum (Vol. 33, No. 1, pp. 10-18). Sage CA: Los Angeles, CA: SAGE Publications.

4. Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. Future healthcare journal, 6(2), 94.

5. Hashmi, H. A. S., & Asif, H. M. (2020). Early detection and assessment of covid-19. Frontiers in medicine, 7, 311.

6. Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. International journal of data mining & knowledge management process, 5(2), 1.

7. Toh, C., & Brody, J. P. (2021). Applications of Machine Learning in Healthcare. Smart Manufacturing: When Artificial Intelligence Meets the Internet of Things, 1-25.

8. Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (Vol. 1, pp. 29-40).

20TH ANNIVERSARY German University in Cairo الجامعة الألمانية بالقاهرة