

FACTORS CONTRIBUTING TO LONG COVID OUTCOMES USING ARTIFICIAL NEURAL NETWORK (ANN)

Nurain Farhanah Yusaini and Izleen Ibrahim
College of Computing, Informatics and Mathematics,
Universiti Teknologi MARA, Perlis Branch
fanaain20@gmail.com and zleen373@uitm.edu.my

ABSTRACT - Background: The COVID-19 pandemic has resulted in a significant number of individuals experiencing long-term symptoms, known as long COVID. Understanding the factors contributing to long COVID outcomes is crucial for effective management and intervention. Objective: This study aimed to identify and rank the factors contributing to long COVID outcomes, develop an Artificial Neural Network (ANN) model, and evaluate its performance in predicting long COVID outcomes. Method: Data from a cross-sectional study by Moy et al. (2022) were used, including variables such as gender, age, BMI, smoking status, comorbidities, and severity of acute COVID-19. The Multilayer Perceptron (MLP) model in IBM SPSS Statistics 27 was employed for data analysis and ANN model development. Findings: Age, smoking status, severity of acute COVID-19, and heart disease emerged as significant factors associated with long COVID outcomes. The developed MLP model achieved an accuracy of 81.3% in predicting long COVID outcomes, with an area under the curve (AUC) of 0.753. Conclusion: This study provides valuable insights into the factors contributing to long COVID outcomes. Age, smoking status, and severity of acute COVID-19 were identified as key predictors of long-term effects. The developed ANN model offers a useful tool for healthcare professionals and policymakers in understanding and addressing the challenges of long COVID. Further research is needed to explore additional variables and validate the findings in diverse populations.

Keywords: Long COVID, factors, Artificial Neural Network, Multilayer Perceptron

1. INTRODUCTION

The study aims to identify the factors contributing to long COVID outcomes using an artificial neural network (ANN) approach. The objectives are to determine the main factors, develop a predictive model, and evaluate the performance of the model in predicting long COVID outcomes. The study focuses on variables such as demographic characteristics, pre-existing health conditions, and severity of the acute COVID-19 infection. The data used is obtained from a research paper by Moy et al. (2022). Limitations include the reliance on self-reported data and the retrospective nature of the study. Despite these limitations, the study is significant as it can provide valuable insights for healthcare resource allocation, inform public health measures and policies, and guide interventions to improve the quality of life for individuals with long COVID. Overall, the study has the potential to contribute to the understanding and management of long COVID and ultimately improve the health and well-being of those affected.

2. METHODOLOGY

This section provides an overview of the data collection process, data description, method of data analysis, and the artificial neural network (ANN) used in the study. The data, obtained from the Plus One website, consists of variables related to factors contributing to long COVID. The ANN approach, specifically the MLP model in IBM SPSS Statistics 27, was employed to analyze the data and develop a predictive model. The model's architecture, activation functions, and training method were specified. Evaluation metrics, including accuracy, precision, recall, and the AUC of the ROC curve, were used to assess the model's performance. The findings of the study will enhance our understanding of long COVID and guide future research and interventions.

3. RESULTS AND DISCUSSION

In this chapter, the findings obtained using the SPSS software are discussed, focusing on data partition, neural network design, model evaluation, and the importance of independent variables. The data partition was performed, with 65.8% of cases assigned to training and 34.2% to testing the ANN models. The neural network design consisted of an input layer with 31 nodes representing various factors, one hidden layer with 4 nodes, and an output layer with 2 nodes for the long COVID outcome. The model's performance was evaluated using metrics such as accuracy, percent incorrect

predictions, and the sum of squares error. Sensitivity analysis was conducted to identify significant input factors for long COVID prediction. The model's evaluation included classification matrices, calculated indicators (accuracy, positive/negative predictive values, recall, specificity, and F1 score), and the area under the curve (AUC) analysis. A predicted versus observed chart was used to assess the model's predictions. The importance of independent variables was analyzed, with age, smoking status, severity of acute COVID-19, and heart disease identified as important factors for predicting long COVID outcomes. These findings provide insights into the predictive performance and variable importance of the ANN model in understanding and addressing long COVID.

4. NOVELTY OF RESEARCH / PRODUCT

The novelty of this research lies in its application of an Artificial Neural Network (ANN) approach to analyze and rank factors contributing to long COVID outcomes. While previous studies have explored the impact of various factors on long COVID, the use of ANN in this context is a novel approach. Additionally, the study provides insights into the relative importance of different factors, such as age, smoking status, and disease severity, which can inform future research and interventions. By utilizing ANN and conducting sensitivity analysis, this research contributes to the originality and advancement of knowledge in understanding long COVID.

5. CONCLUSION

In conclusion, this study utilized an Artificial Neural Network (ANN) approach to identify and rank factors contributing to long COVID outcomes. Age, smoking status, severity of acute COVID-19, and heart disease were found to be significant factors. The ANN model showed promising performance in predicting long COVID outcomes, and recommendations were made for future research to expand the dataset, include additional variables, and conduct intervention studies for improving long COVID outcomes.

REFERENCES

- Alwan, N. A. (2020). Surveillance is underestimating the burden of the COVID-19 pandemic. *The Lancet*, 396(10252), Article e24. [https://doi.org/10.1016/S0140-6736\(20\)31823-7](https://doi.org/10.1016/S0140-6736(20)31823-7)
- Moy, F. M., Hairi, N. N., Lim, E. R. J., & Bulgiba, A. (2022). Long COVID and its associated factors among COVID survivors in the community from a middle-income country-An online crosssectional study. *PLoS ONE*, 17(8). <https://doi.org/10.1371/journal.pone.0273364>
- Subramanian, A., Nirantharakumar, K., Hughes, S., Myles, P., Williams, T., Gokhale, K. M., Taverner, T., Chandan, J. S., Brown, K., Simms-Williams, N., Shah, A. D., Singh, M., Kidy, F., Okoth, K., Hotham, R., Bashir, N., Cockburn, N., Lee, S. I., Turner, G. M., Haroon, S. (2022). Symptoms and risk factors for long COVID in non-hospitalized adults. *Nature Medicine*, 28(8), 1706–1714. <https://doi.org/10.1038/s41591-022-01909-w>