Appl. Math. Inf. Sci. **19**, No. 1, 131-139 (2025)

Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/190111

Artificial Intelligence models versus a machine learning aided model for severity prediction of COVID-19: A Population-Based Study

Eman Khorsheed [1,*] *and Hawra Alshowaikh*[2]

[1]Department of Mathematics, College of Science, University of Bahrain, Bahrain
[2]Faculty of Medicine Kasr Al-Ainy, Cairo University, Cairo, Egypt

Received: 2 Feb.. 2024, Revised: 27 May 2024, Accepted: 5 Nov. 2024

Abstract: The innovative potentials of the Artificial Neural Network (ANN) and the Random Forest artificial intelligence technologies along with a machine learning aided logistic regression technique were utilized to predict COVID-19 severity. The three models were compared based on sensitivity, specificity, and overall model accuracy. Data of 205 patients were analyzed in this study. 70.6% of the observations were used for model training and 29.4% for model validation. In each approach, ten potential risk factors for severity symptoms were considered and evaluated. At the training phase, the overall accuracy of the ANN, logistic regression, and Random Forest models were 84.7%, 83.3%, and 81.4%, respectively. However, the ANN model outperformed the other two models during the validation process and scored an overall accuracy of 85.0% versus 71.7%. This superior performance underscores the robustness of the ANN approach in the context of predictive accuracy for COVID-19 severity.

Keywords: Artificial Neural Network, Logistic Regression, Random Forest, prediction, pandemic

1 Introduction

Coronaviruses are a class of RNA viruses of different types, some of which cause disease in humans, animals and birds. On 12 January 2020, the World Health Organization (WHO) [1] confirmed the existence of a new strain of coronavirus, initially named novel coronavirus 2019. In February 2020, WHO named this group of viruses COVID-19. The outbreak was first detected in December 2019 in the Chinese city of Wuhan [2]. COVID-19 spread rapidly around the world early 2020, but was declared a global pandemic on 11 March 2020.

The novel coronavirus disease (Covid-19) is classified as a respiratory infection that spreads between people through direct or indirect contact with infected individuals. In their work, Hattaf et al. [3] developed a mathematical model that can be used to describe the dynamics of coronavirus by considering the different modes of disease transmission and the influence of pro-social behavior. The impact of the disease on individuals varies widely. After 2-14 days of exposure to the virus, patients may develop mild symptoms to severe disorders. According to Public Health England [4], 85% of individuals infected with Covid-19 initially exhibit at least one of the following primary symptoms: fever or chills, fatigue, and dry cough. These three symptoms are commonly observed in clinical settings. Additional symptoms may include a sore throat, anosmia (loss sense of smell), dysgeusia (distortion or loss sense of taste), headache, body aches, diarrhea, nausea, chest pain, and breath difficulties. Furthermore, a multicenter descriptive study in China found that 80% of confirmed COVID-19 cases experienced mild to moderate symptoms [5].

As of March 26, 2021, the total global number of confirmed COVID-19 cases reached 125,436,393 with 2,756,768 deaths [6]. The countries most severely impacted by the pandamic include the US, Brazil, Mexico, India, Russia, the UK, France, and Italy. There have been limited studies aimed at predicting the short-term and long-term trajectories of the novel coronavirus, particularly in these profoundly heavily affected nations. For further reference, see studies [7,8].

The Covid-19 pandemic in Bahrain is part of the broader global crisis, with the first case confirmed by the

---

[*] Corresponding author e-mail: ekhorsheed@uob.edu.bh

Bahrain Ministry of Health on 21 February 2020 [9]. As of 26 March 2021, the Ministry reported 139,124 confirmed cases, with 512 deaths due to the of virus. In January 2021, Bahrain introduced mobile vaccination units aimed at immunizing the elderly population, positioning it as the third most vaccinated country globally at that time. Bahrain also continued to offer the COVID-19 vaccine free of charge to all of its citizens and residents. By 26 March 2021, a total of 243,940 individuals in Bahrain were fully vaccinated. In the week leading up to 26 March 2021, an average of 763 new confirmed cases were reported. This number represented 325 cases per 100,000 people, which is equivalent to 96% of the peak recorded on 16 February 2021, as reported by Reuters [10]. These statistics marked the Kingdom of Bahrain as one of the most heavily impacted countries in the Middle East during the period leading up to 1 April 2021 [11].

During the COVID-19 pandemic, the early identification of patients at high-risk of severe outcomes is crucial [12], not only for delineating the clinical and epidemiological characteristics of the disease but also for implementing optimized treatment protocols. Such measures are essential for delivering high-quality medical services and achieving better healthcare outcomes. Despite the urgency , only a few studies have aimed at achieving this goal.

Bhandari et al. [13], for example, developed a logistic regression model to predict COVID-19 mortality. Their model incorporated factors are age, sex, symptom presence, and routine hematologic measurements from laboratory-confirmed cases. Among the 21 regressors analyzed, the study highlighted the differential neutrophil count and random blood sugar (RBS) levels as significant predictors. These findings suggest that both factors can serve as rapid, reliable indicators for assessing the risk of mortality from the coronavirus, thereby facilitating quicker and more effective clinical decision-making.

Agbehadji et al. [14] explored the integration of big data recent platforms, artificial intelligence (AI) models, and nature-inspired algorithms in the detection of Covid-19 cases and contact tracing of the disease. This work highlights the potential of integrating advanced computational technologies to enhance the accuracy and efficiency of tracking and managing the novel coronavirus.

In a related vein, Albahri et al. [15] conducted a systematic review of state-of-the art methodologies and techniques in developing coronavirus forecasting models. Their review focused on the utilization of data mining and machine learning (ML) tools. This comprehensive study helps identify the strengths and limitations of current approaches and suggests directions for future research in improving predictive models for Covid-19 and potentially other emerging infectious diseases.

The identification and evaluation of potential severe Covid-19 risk factors is of paramount importance for guiding health sectors to provide the best medical care, improve infection outcomes, and allocate appropriate resources throughout the disease's progression. Globally, the correlation between the severity of Covid-19 and potential influencing factors continue to be a significant area of research.

The literature indicates that certain demographic characteristics are strongly associated with an increased severity of COVID-19. Notably, older age is consistently identified as a key risk factor [16] and [17].

Furthermore, some studies have indicated that male sex may also be a significant risk factor, suggesting that being male may independently correlate with a more complicated course of COVID-19, including mortality [18]. This observation has prompted further investigation into biological and behavioral factors that might contribute to this disparity, such as differences in immune system responses, lifestyle choices, and pre-existing health conditions that vary between genders.

In addition, comorbidities such as cardiovascular disease, immunosuppression, chronic kidney and lung diseases, diabetes mellitus, hypertension, active cancer, certain blood disorders, and obesity may predispose patients with Covid-19 to an elevated risk of intubation and death [19]. For example, the Disease Control and Prevention centers in U.S. [19] consider obesity, sickle cell anemia, and asthma, as risk factors for severe coronavirus outcomes. Despite the U.S. reports, the existing literature lacks sufficient population-based studies that support these determinations until this stage of the epidemic's spread.

The Ministry of Health of Bahrain has indicated that the Kingdom of Bahrain is one of the countries most affected by diabetes, with incidence rate of approximately 14.7% among the adult population. Additionally, genetic blood disorders are notably prevalent in Bahrain. These include sickle cell anemia (SCD), Glucose-6-phosphate dehydrogenase deficiency (G6PD), and thalassemia [20].

Given the substantial burden of these comorbidities diseases and the lack of population-based studies supporting their impact on Covid-19 severity, the primary objective of this research was to develop and validate AI and ML-driven population-based models. These models aim to predict severe COVID-19 cases using a set of potential demographic and clinical characteristics. Another crucial goal was to compare the effectiveness of the developed models in terms of accuracy, sensitivity, and specificity.

Furthermore, this study conducted a sensitivity analysis on the developed classifiers in order to gain insight into the relative importance of each considered potential risk factor in predicting patients at high risk of severe Covid-19. This analytical approach not only enhances understanding of the contribution of comorbidities and demographic factors to the severity of Covid-19, but also aids in the tailoring of interventions and the optimization of resource allocation for those most at risk in similar settings globally.

2 METHODOLOGY

2.1 COVID-19 Data and Descriptive Statistics

The study focuses on a sample of 205 individuals who contracted COVID-19 between 15 March 2020 and 31 March 2021 in the Kingdom of Bahrain. Due to the limited availability of detailed patient information on COVID-19 infections and deaths, demographic and clinical characteristics, data were primarily gathered through an online questionnaire distributed on patients or relatives of those deceased from COVID-19. The questionnaire was explicitly stated to have a specific research purpose in order to ensure that respondents were aware of the context of their participation. Additional data sources included the Bahrain Ministry of Health [10], Our World in Data [21], the World Health Organization (WHO) [22], and the Johns Hopkins University Coronavirus Resource Center [23].

The study's descriptive statistics revealed significant demographic details about the infected cohort. The sample comprised 85.4% Bahraini residents and 14.6% foreign residents. The average age of the participants was 40 years. In terms of age distribution, 40.5% of infected individuals were aged between 15 and 29 years, 43.9% fell within the 30 to 59 years age range, and 15.7% were older than 60 years but younger than 90 years. Table 1 displays a detailed description of the participant characteristics.

2.2 Severity potential predictors and response variable

The present study examined ten demographic and clinical risk factors that could potentially lead to severe Covid-19 illness. The demographic factors included gender and age, while the clinical factors encompassed obesity, diabetes mellitus, asthma, cardiovascular disease, neurological disorders, cancer, hypertension, and genetic blood disorders. The age of a participant is the only continuous variable, while all other variables are binary. For instance, a medical condition is noted as present (1) or absent (0), and gender is coded as male (1) or female (0).

In this study, the severity of COVID-19 is defined based on the following clinical status: 1- hospitalized patients due to, e.g., acute dyspnea, low oxygen saturation level, the need for mechanical ventilation, or admission to the intensive care unit (ICU), and 2- death. The response variable, COVID-19 severity, is categorized into two levels: non-severe (0) and severe (1). While a proportion of infected individuals experienced only mild disease, 126 cases (61.5%) encountered complicated symptoms of COVID-19 of which 40 (19.5%) patients could not survive.

Chi-square test was conducted using the statistical software SPSS to determine the significant relationships between the considered factors and the dependent variable of COVID-19 severity. Most factors were found to be significantly associated with severity of Covid-19 (p-value less than 0.05), with the exceptions of gender and neurological disorder, as shown in Table 2.

3 COVID-19 SEVERITY RISK PREDICTION MODELS

Despite the Chi-square test indicating that gender and neurological disorders were not statistically significant factors in severe COVID-19 illness, as detailed in Table 2, this finding contrasts with several prior studies. For example, a meta-analysis cited in [24] identified male gender as a potential risk factor for developing severe symptoms of the coronavirus. This discrepancy prompted a comprehensive re-evaluation of all ten demographic and clinical characteristics of patients with COVID-19, using advanced AI and machine learning (ML) technologies to reassess the associations between these potential predictors and the severity of the coronavirus.

In this research, SPSS was employed to build an Artificial Neural Network (ANN) model with the objective of predicting patients at high risk of developing severe forms of the novel coronavirus disease (COVID-19). For comparative analysis, a Logistic Regression (LR) model and a Random Forest (RF) model were also developed using the machine learning capabilities of the statistical software R.

Table 1: Descriptive Statistics of COVID-19 Data.

Factor	Percentage
Gender (Male)	47.8%
Diabetes	19%
Hypertension	22%
Obesity	31.2%
Asthma	19.5%
Cancer	4.9%
Cardiovascular Disease	10.7%
Neurological Disorder	2%
Genetic Blood Disorder	24.4%

This multi-model approach allows for a robust analysis across different methodologies, enhancing the reliability of the predictions regarding which factors are truly indicative of severe COVID-19 cases.

In this study, the COVID-19 dataset was split into two subsets to facilitate the training and testing of the developed models. The training set comprised 70.6% of the observations (145 patients), while the testing set included 29.4% (60 patients). The methodologies employed for each model are presented below in detail.

3.1 Artificial Neural Network COVID-19 Prediction Model

The integration of Artificial intelligence (AI) into the medical sector has become increasingly prevalent over the past decade. This is particularly evident in the development of complex and non-linear models that can handle a wide array of clinical data [25]. Recently, there has been growing evidence that AI can play a role in combating the ongoing COVID-19 pandemic at both the individual and societal levels [26]. One common application of AI in this field is classification, a type of data mining that involves the assignment of new observations within a dataset to predetermined categories based on patterns learned from a training set with known classes [27].

A non-linear artificial neural network (ANN) prediction classification model was constructed here by configuring multiple layers that simulate the way neurons operate in the human brain. The network was optimized to identify complex patterns and interactions among the input variables using the multi-layer perceptron (MLP) algorithm.

3.2 Logistic Regression COVID-19 Prediction Model

Logistic regression is a statistical method that is widely used for binary classification. It is particularly effective for predicting a binary response variable from a set of continuous or categorical data. This regression technique is notable for its proficiency in dealing with small datasets that may contain non-linearly separable data. This is a common scenario in many medical contexts, including disease prediction and health studies. It has been employed extensively in various domains, including the prediction of the onset of epidemics, the management of chronic diseases, and the assessment of cancer survivability [28,29].

In logistic regression, the model estimates the probability that a binary dependent variable equals 1 (for instance, indicating that an individual has contracted a disease) versus 0 (indicating no disease). This is achieved by relating the log-odds of the dependent variable to a linear combination of the independent variables or risk factors [30]. The method outputs probabilities that are subsequently mapped to binary categories based on a decision threshold, typically 0.5, rendering it a practical tool for decision-making in clinical settings.

In the context of this research, a logistic regression model was developed using the machine learning tools available in the R statistical software. The aim was to predict the risk of severe COVID-19 symptoms, enabling health care providers to identify high-risk patients at an early stage.

Table 2: Chi-Square Test Results.

Factor	p-value
Age	< 0.001
Gender (Male)	0.865
Diabetes	< 0.001
Hypertension	< 0.001
Obesity	< 0.001
Asthma	< 0.001
Cancer	0.015
Cardiovascular Disease	< 0.001
Neurological Disorder	0.249
Genetic Blood Disorder	0.249

3.3 Random Forest COVID-19 Prediction Model

Decision trees represent a type of supervised learning algorithm that can be employed for both regression and classification tasks. Their recursive splitting of observations into branches results in a tree-like structure, with each node representing a decision point based on specific conditions. The selection of the optimal attribute to split the data at each node is based on a specific criterion, such as Gini impurity or entropy in classification, and variance reduction in regression. ML algorithms developed in [31] and [32] are used to seek decision tree predictive models.

The Random Forest technique is an extension of the decision tree method that incorporates ensemble learning. It operates by creating multiple decision trees at the training stage to output an uncorrelated forest of trees that represents the mode of the underlying individual decision trees. Random Forest has the advantage of being insensitive to over-fitting, a well-known problem associated with decision trees. Thus, while individual decision trees provide straightforward, interpretable models that are easy to implement, Random Forest leverages the power of multiple trees to enhance predictive accuracy and generalization ability across unseen data [33]. This makes ensemble models like Random Forest more suitable for applications requiring high reliability and precision in predictions, such as in complex disease prediction scenarios, customer behavior analysis, and other areas requiring robust analytical solutions.

In this study, the Random Forest model was employed to predict the severity of COVID-19, utilizing an ensemble of 500 decision trees to enhance predictive accuracy and stability. Each tree in the forest was permitted to consider three randomly selected predictors as potential candidates at each split. This approach not only helps to reduce the correlation between individual trees in the forest, thereby increasing the ensemble's robustness, but also aids in overcoming the overfitting issue commonly seen with single decision trees.

The dataset was divided into two distinct subsets: 145 randomly selected observations were utilized for the training of the model, while the remaining 60 observations were set aside for the purpose of validation. This division of the data ensures that the model can be trained on a substantial portion of the data while still having a separate dataset to evaluate its performance. The training set enables the Random Forest algorithm to learn the relationships and patterns between the predictors and the outcome variable. In contrast, the validation set tests the model's ability to generalize these patterns to new, unseen data.

Using 500 trees is a strategic choice that balances between too few trees, which might not capture sufficient complexity, and too many, which could unnecessarily extend the training time without a proportionate gain in performance. By using multiple trees and averaging their predictions, the Random Forest model typically achieves high accuracy and is less sensitive to outliers and noise within the dataset.

4 Results

In this study, an Artificial Neural Network (ANN) model was developed to predict the severity of cases of the novel coronavirus disease (COVID-19). This ANN model was then compared with two other predictive models: Logistic Regression (LR) and Random Forest (RF). Each of these models represents a distinct approach to classification and prediction tasks, leveraging different aspects of machine learning theory and application. Predictive performance of each classifier was assessed based on accuracy, sensitivity and specificity.

4.1 Prediction Evaluation

The evaluation of binary classification models is based on two key metrics: sensitivity and specificity. Sensitivity measures the proportion of positive cases that are correctly identified, while specificity assesses the rate of negative cases that are correctly classified. Accuracy is a measure of the overall correctness of the model. It is the ratio of true predictions (both true positives and true negatives) to the total number of cases examined. The prediction results based on the training and testing subsets are presented in the form of confusion matrices, which are displayed in Table 3. As shown in this table, the severity of COVID-19 was divided into two classes: mild and severe. The ANN model achieved an overall prediction accuracy of 84.7% based on the training dataset, with a sensitivity of 85.9% and a specificity of 83.1%. The overall prediction accuracy of the LR model was found to be 83.3%, with a sensitivity of 83.5% and a specificity of 83.1%. The RF model achieved an overall prediction accuracy of 81.4%, with a sensitivity of 83.2% and a specificity of 78.6%. In the testing subset, the ANN model achieved an overall prediction accuracy of 85.0%, with a sensitivity of 87.8% and a specificity of 78.9%. In contrast, the LR model produced predictions with an overall accuracy of approximately 71.7%, and a sensitivity of 80.5%, but a much lower specificity of 52.6%. The RF model achieved an overall prediction accuracy of 71.7%, a sensitivity of 64.9%, and a specificity of 82.6%. A comparison of the classification results derived from the testing phase with those obtained during the training phase revealed a notable decline in the prediction accuracy of the logistic regression (LR) model for mild cases, while the accuracy of the severe cases remained largely unaffected.

In contrast to the LR model, the RF prediction accuracy of the mild cases increased, whereas that of the severe cases decreased. These changes in performance have reduced the overall prediction accuracy of both classifiers, namely the LR and the RF, by approximately 10.0% each. However, the overall performance of these two models remained relatively good with an identical accuracy score of 71.7%. Conversely, the ANN model's predictions remained stable and achieved an overall accuracy of approximately 85.0%. Consequently, the ANN is considered the most effective model among those developed in this study.

Sensitivity analysis is a critical tool in the realm of statistical modeling and machine learning, especially in contexts where it is crucial to understand the impact of various input variables on the output of a prediction model. It has been widely used to identify the degree to which each variable contributes to the identification of each class of the response variable.

In this work, a sensitivity analysis was conducted across different modelling frameworks, namely the Artificial Neural Network (ANN), Logistic Regression, and Random Forest (RF).

Sensitivity analysis in ANN involves perturbing each input factor slightly and observing the change in the output variable. This technique identifies which factors have the most substantial impact on the model's predictions. It can reveal complex nonlinear relationships and interactions among variables that are typical in neural network models.

In Random Forest, sensitivity analysis typically involves evaluating the decrease in model accuracy when a specific factor is omitted or permuted. This measure, often referred to as the ”importance” of a factor, which indicates how vital a particular predictor is for making accurate predictions. The RF model naturally lends itself to this kind of analysis because each tree in the forest provides a natural variance estimate that can be used to assess factor importance.

In contrast, the importance of variables in logistic regression is commonly assessed using the z-statistic, which measures the influence of each predictor while holding other variables constant. The absolute value of the z-statistic for each predictor provides a scale of

Table 3: Performance of the ANN, LR, and RF Models for COVID-19 Severity.

ANN Model				LR Model				RF Model			
Train Data											
Sample	Mild	Severe	Correct	Sample	Mild	Severe	Correct	Sample	Mild	Severe	Correct
Train Mild	49	10	83.1%	Train Mild	49	10	83.1%	Train Mild	44	12	78.6%
Train Severe	12	73	85.9%	Train Severe	14	71	83.5%	Train Severe	15	74	83.2%
Overall %	42.4%	57.6%	84.7%	Overall %	43.8%	56.2%	83.3%	Overall %	40.7%	59.3%	81.4%
Test Data											
Test Mild	15	4	78.9%	Test Mild	10	9	52.6%	Test Mild	19	4	82.6%
Test Severe	5	36	87.8%	Test Severe	8	33	80.5%	Test Severe	13	24	64.9%
Overall %	33.3%	66.7%	85.0%	Overall %	30.0%	70.0%	71.7%	Overall %	53.3%	46.7%	71.7%

importance, with higher values indicating a greater impact on the model's outcome. This method, described in [34], offers a straightforward measure of variable importance.

Fig. 1 and Table 4 present the sensitivity analysis performed in this research across all three COVID-19 severity prediction models: ANN, RF, and Logistic Regression.

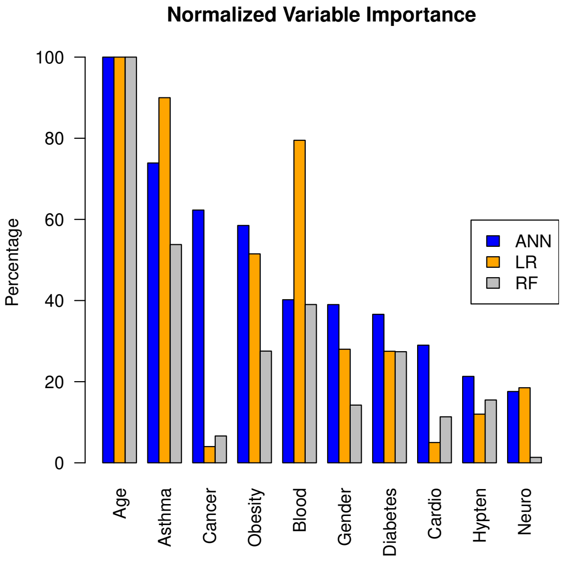

Fig. 1: Normalized Factor Importance of the 10 considered predictors for the ANN, LR, and the RF models.

The results of the conducted sensitivity analysis displayed in Table 4 revealed that older age- a factor consistently identified in several previous studies [35,36, 37]- along with active asthma, and subsequently cancer (in the ANN approach) or genetic blood disorders (in the LR and RF approaches), followed by obesity, have highly dominated the predictions of the ANN, LR, and RF classifiers with total factor importance scores of 61.5%, 77.9%, and 74.3%, respectively. Although the LR model considered the group of patients with blood disorders at elevated risk of Covid-19 severity, the normalized variable importance percentages in Fig.1 moderately ranked this factor in the AI approaches, namely the ANN and RF. The later results are in line with some available studies on severe COVID-19 symptoms in SCD patients [38,39], and several anecdotal reports [40,41,42,43].

Male gender appeared to have more influence on aggravating the severe symptoms of the infection than the female sex. This result is consistent with several other studies available in literature such as [2,43]. The pairwise Chi-square test conducted initially has suggested significant associations between the dependent variable of COVID-19 severity and three other diseases, namely diabetes mellitus, cardiovascular disease, and hypertension. According to the generated sensitivity analysis of the new models, these diseases have revealed low to moderate influence on the development of Covid-19 severe outcomes. The current results are supported by some existing studies presented in [43,45, 46,47]. These unexpected relationships can be attributed to the prevalent and strong associations between such diseases and aging, where older age was ranked as the most influential factor on COVID-19 severity by all three models. Our conclusion is in line with the findings of the meta-analysis presented in [49].

Finally, a factor that was found to be weakly significant in predicting COVID-19 by all approaches is the neurological disorder. However, this result may be dominated by the small portion of participants affected by the disease that were involved in the study.

5 Conclusion

This research makes a substantial contribution by comparing the performance of two AI models (Artificial Neural Network and Random Forest) and a traditional statistical model (Logistic Regression) in predicting COVID-19 severity. The study classified COVID-19 cases as mild (0) or severe (1), including death. These models were proposed to analyze COVID-19 patient data, predict disease severity/death in relation to ten potential risk factors, and gain insight into the relative influence of each factor on the risk of adverse COVID-19 complications. The predictors included were sex, age, obesity, diabetes mellitus, asthma, cardiovascular disease, neurological

Table 4: Factor Importance & ranking of the 10 predictors for the ANN, LR, and RF models.

ANN	Ranking	LR	Ranking	RF	Ranking
Age (20.9)	1	Age (24.3)	1	Age (33.7)	1
Asthma (15.4)	2	Asthma (21.8)	2	Asthma (18.1)	2
Cancer (13.0)	3	Blood disorder (19.3)	3	Cancer (13.2)	3
Obesity (12.2)	4	Obesity (12.5)	4	Obesity (9.28)	4
Blood disorder	5	Gender (6.8)	5	Diabetes (9.23)	5
Gender (8.2)	6	Diabetes (6.7)	6	Hypertension (5.2)	6
Diabetes (7.7)	7	Neurological disorder (4.5)	7	Gender (4.8)	7
Cardiovascular disease (6.0)	8	Hypertension (2.8)	8	Cardiovascular disease (3.8)	8
Hypertension (4.5)	9	Cardiovascular disease (1.2)	9	Cancer (2.2)	9
Neurological disorder (3.7)	10	Cancer (0.1)	10	Neurological disorder (0.45)	10

disorders, cancer, hypertension, and genetic blood disorders. The 205 COVID-19 cases analyzed were reported in Bahrain from March 15, 2020, to March 31, 2021.

The ANN model was developed using the SPSS software, while the LR and the RF models were built and tested using ML tools available in R. At the testing phase, the ANN model preformed best among the three estimated models and achieved the highest classification accuracy of 85.0%. Both the LR and the RF models achieved an overall prediction accuracy of 71.7% each. This superior performance of the ANN model suggests its potential suitability for more complex data patterns often present in medical datasets.

In addition to predicting Covid-19 severity, sensitivity analysis was performed on all developed models to gain insight into the relative importance of the potential risk factors in predicting Covid-19 severity. The results showed that older age was by far the most important risk factor for severity.

Sensitivity analysis results revealed older age as the most critical risk factor across all models, aligning with global observations about the virus's impact on older populations.

Asthma, not traditionally considered a major risk factor in earlier studies, emerged as the second most significant predictor in this research. This finding supports the recent study in [50], which considered patients with asthma within 1 year prior to COVID-19 as a group at increased risk for severe disease.

Cancer and genetic blood disorders were also significant predictors but varied in their impact across the different models. Males were found to be more likely to develop severe disease than females, but gender was not as predictive of severity. Neurological disorders were found to contribute minimally to severe outcomes, which could be due to the limited cases within the considered sample.

With the available innovative AI technologies, COVID-19 patients at high risk of progressing to severe infection can be identified early for targeted treatment and rapid medical care. In addition, healthcare professionals can take rapid and appropriate action to better control the severe infection and manage available resources more efficiently to improve the quality of care provided and outcomes.

The effectiveness of the ANN model in this study provides a strong case for its use in similar predictive modeling scenarios, encouraging further exploration and refinement of AI techniques in healthcare.

In summary, this research not only demonstrates the utility of advanced AI models in managing health crises like the COVID-19 pandemic but also enriches the existing literature by highlighting novel insights into risk factor prioritization and model selection for predictive accuracy.

References

[1] World Health Organization. Novel Coronavirus- China. Disease Outbreak news. https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/.

[2] H. Peckham, N. de Gruijter, C. Raine, A. Radziszewska, C. Ciurtin, L. Wedderburn, E. Rosser, K. Webb, and C. Deakin, Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission, Nat. Commun., **11(1):6317** (2020). https://doi.org/10.1038/s41467-020-19741-6.

[3] K. Hattaf, A. Mohsen, J. Harraq, and N. Achtaich, Modelling the dynamics of Covid-19 with carrier effect and environmental contamination, International Journal of Modeling, Simulation, and Scientific Computing, **12(3):2150048** (2021).

[4] Public Health England, https://www.gov.uk/coronavirus.

[5] J. Wu, J. Liu, X. Zhoa, C. Liu, W. Wang, D. Wang, W. Xu, C. Zhang, J. Yu, B. Jiang, H. Cao, and L. Li, Clinical characteristics of imported cases of coronavirus disease 2019 (COVID-19) in Jiangsu Province: amulticenter descriptive study, Clin Infect Dis, **71(15)**, 706-712 (2020).

[6] Worldometers, https://www.worldometers.info/coronavirus/.

[7] T. Chakraborty and I. Ghosh, Real-time forecasts and risk assessment of novel coronavirus(COVID-19) cases: A

data-driven analysis, Chaos Solitons Fractals, **135:109850** (2020).

[8] I. Ghosh and T. Chakraborty, An integrated deterministic?stochastic approach for forecasting the long-term trajectories of COVID-19, International Journal of Modeling, Simulation,and Scientific Computing, **12(3):2141001** (2021).

[9] Bahrain Ministry of Health. https://www.moh.gov.bh/COVID19/.

[10] Reuters COVID-19 Tracker. https://graphics.reuters.com/world-coronavirus-tracker-and-maps/countries-and-territories/bahrain/.

[11] BBC News. https://www.bbc.com/news/world-51235105.

[12] J. Jin, P. Bai and W. He, , Gender Differences in Patients with COVID-19: Focus on Severity and Mortality, Frontiers in Public Health, **8:152** (2020).

[13] S. Bhandari, S. Shaktawat and A. Tak, Logistic regression analysis to predict mortality risk in COVID-19 patients from routine hematologic parameters, Ibnosina Journal of Medicine & Biomedical Sciences, **12(2)**, 123-129 (2020).

[14] I. Agbehadji, B. Awuzie, A. Ngowi, and R. Millham, Review of Big Data Analytics, Artificial Intelligence and Nature-Inspired Computing Models towards Accurate Detection of COVID-19 Pandemic Cases and Contact Tracing, International Journal of Environmental Research and Public Health, **17(15), 5330** (2020).

[15] A. Albahri, R. Hamid, J. Alwan, Z. Al-qays, A. Zaidan, B. Zaidan, A. Albahri, A. AlAmoodi, J. Khlaf, E. Almahdi, E. Thabet, S. Hadi, K. Mohammed, M. Alsalem, J. Al-Obaidi, and H. Madhloom, Role of biological Data Mining and Machine Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19): A Systematic Review, Journal of Medical Systems, **44(7):122** (2020).

[16] M. Cecconi, D. Piovani, E. Brunetta, A. Aghemo, M. Greco, M. Ciccarelli, C. Angelini, A. Voza, P. Omodei, E. Vespa, N. Pugliese, T. Parigi, M. Falci, S. Danese and S. Bonovas, Early predictors of clinical deterioration in a cohort of 239 patients hospitalized for Covid-19 infection in Lombardy, Italy, Journal of Clinical Medicine Research, **9(5):1548** (2020).

[17] J. Galloway, S. Norton and R. Barker, A clinical risk score to identify patients with Covid-19 at high risk of critical care admission or death: an observational cohort study, Journal of Infections, **81(2)**, 282-288 (2020).

[18] X. Li, S. Xu, M. Yu, K. Wang, Y. Tao, Y. Zhou, J. Shi, M. Zhou, B. Wu, Z. Yang, C. Zhang, J. Yue, Z. Zhang, H. Renz, X. Liu, J. Xie, M. Xie and J. Zhao, Risk factors for severity and mortality in adult Covid-19 inpatients in Wuhan, Journal of Allergy and Clinical Immunology, **146(1)**, 110-118 (2020).

19 B. Gallo Marin, G. Aghagoli, K. Lavine, L. Yang, E. Siff, S. Chiang, T. Salazar-Mather, L. Dumenco, M. Savaria, S. Aung, T. Flanigan and I. Michelow, Predictors of Covid-19 severity: A literature review, Reviews in Medical Virology, **31(1)**, 1-10 (2021).

[19] Centers for Disease Control and Prevention (CDC). People with Certain Medical Conditions, https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html.

[20] S. Al Arrayed, Campaign to control genetic blood diseases in Bahrain, Community Genet, **8(1)**, 52-55 (2005).

[21] Our World in Data, https://ourworldindata.org/coronavirus.

[22] World Health Organization. Director-General?s remarks at the media briefing on 2019-nCoV on 11 February 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020.

[23] Coronavirus Resource Center at John Hopkins University. https://coronavirus.jhu.edu/map.html.

[24] C. Huang, Y. Wang and X. Li, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, Lancet., **395**, 497-506 (2020).

[25] T. Guggenberger, J. Lockl, M. Rogllnger, V. Schlatt, J. Sedlmeir, J. Stoetzer, N. Urbach, F. Volter, Emerging Digital Technologies to Combat Future Crises: Learning from COVID-19 to be prepared for the future, International Journal of Innovation and Technology Management, **18(4):2140002** (2021).

[26] S. Alkheder, M. Taamneh, S. Taamneh, Severity Prediction of Traffic Accident Using an Artificial Neural Network, J. Forecast, **36**, 100-108 (2017).

[27] S. Nusinovici, Y. Tham, M. Yan, D. Ting, J. Li, C. Sabanayagam, T. Wong and C. Cheng, Logistic regression was as good as machine learning for predicting major chronic diseases, Journal of Clinical Epidemiology, **122**, 56-69 (2020).

[28] D. Delen, G. Walker and A. Kadam, Predicting breast cancer survivability: a comparison of three data mining methods, Artif Intell Med, **34**, 113-127 (2005).

[29] C. Chang, C. Wang, and B. Jiang, Using data mining techniques for multi-diseases prediction modeling of hypertension and hyperlipidemia by common risk factors, Expert Syst Appl, **38**, 5507-5513 (2011).

[30] L. Breiman, J. Friedman, R. Olshen and C. Stone, Classification and regression trees, Monterey, CA:Wadsworth & Brooks/Cole Advanced Books & Software, 1984.

[31] J. Quilan, Induction of decision trees, Machine Learning, 81-106 (1986).

[32] C. Dahinden, An improved Random Forests approach with application to the performance prediction challenge datasets. Hands on Pattern Recognition. Microtome, 2009.

[33] M. Kuhn, Building predictive models in R using the caret package, J Stat Softw, **28**, 1-26 (2008).

[34] Z. Zheng, F. Peng, B. Xu, J. Zhao, H. Liu, J. Peng, Q. Li, C. Jiang, Y. Zhou, S. Liu, C. Ye, P. Zhang, Y. Xing, H. Guo and W. Tang, Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis, J Infect., **81(2)**, 16-25 (2020).

[35] T. Johannesen, S. Smeland, S. Aaserud, E. Buanes, A. Skog, G. Ursin and A. Helland, COVID-19 in Cancer Patients, Risk Factors for Disease nd Adverse Outcome, a Population-Based Study from Norway, Frontiers in Oncology, **11:652535**, 2021.

[36] A. Rahman and N. Sathi, Risk factors of the severity of COVID-19: a Meta-Analysis, Int J Clin Pract, **75(7):e13916**, 2021.

[37] A. Clift, D. Saatci, C. Coupland, H. Dambha-Miller and J. Hippisley-Cox, Sickle Cell Disorders and Severe COVID-19 Outcomes: A Cohort Study, Annals of Internal Medicine, **174(10)**, 1483-1487 (2021).

[38] N. Roy, P. Telfer, P. Eleftheriou, J. de la Fuente, E. Drasar, F. Shah, D. Roberts, W. Atoyebi, S. Trompeter, D. Layton, S. Lugthart, S. Stuart-Smith, S. Chakravorty, J. Wright, J. Porter, B. Inusa, and J. Howard. Protecting vulnerable patients with inherited anaemias from unnecessary death during the COVID-19 pandemic, British J Haematol, **189(4)**, 635-639 (2020).

[39] F. Beerkens, M. John, B. Puliafito, V. Corbett, C. Edwards and D. Tremblay, Covid-19 pneumonia as a cause of acute chest syndrome in an adult sickle cell patient, American Journal of Hematology, **95(7)**, 154-156 (2020).

[40] K. McCloskey, J. Meenan, R. Hall, and D. Tsitsikas, COVID-19 infection and sickle cell disease: a UK centre experience, British J Haematol, **190(2)**, 57-58 (2020).

[41] G. De Luna, A. Habibi, J. Deux, M. Colard, A. Pham, Hung d'Alexandry d'Orengiani, F. Schlemmer, N. Joher, C. Kassasseya, J. Michel, C. Ourghanlian, M. Mihel, A. Dessap, and P. Bartolucci, Rapid and severe COVID-19 pneumonia with severe acute chest syndrome in a sickle cell patient successfully treated with tocilizumab, American Journal of Hematology, **95(7)**, 876-878 (2020).

[42] F. Hussain, F. Njoku, S. Saraf, R. Molokie, V. Gordeuk and J. Han, COVID-19 infection in patients with sickle cell disease, British J Haematol, **189(5)**, 851-852 (2020).

[43] G. Bwire, Coronavirus:Why Men are More Vulnerable to COVID-19 Than Women? SN Comprehnsive Clinical Medicine, **2**, 874-876 (2020).

[44] M. Rodacki, Severity of COVID-19 and diabetes mellitus: there is still a lot to be learned, Arch Endocrinol Metab, **64**, 195-196 (2020).

[45] J. Li, X. He, Y. Yuan, W. Zhang, X. Li, Y. Zhang, S. Li, C. Guan, Z. Gao and G. Dong, Meta-analysis investigating the relationship between clinical features, outcomes, and severity of severe acute respiratory syndrome coronavirus 2 (SARS-Cov-2)pneumonia, American Journal of Infection Control, **49**, 82-89 (2020).

[46] G. Lippi, J. Wong and B. Henry, Hypertension in patients with coronavirus disease 2019 (Covid-19): a pooled analysis, Pol Arch Intern Med, **130**, 304-309 (2020).

[47] H. Liu, S. Chen, M. Liu, H. Nie and H. Lu, Comorbid Chronic Diseases are strongly correlated with disease severity among COVID-19 patients: A systematic review and meta-analysis, Aging and Disease, **11(3)**, 668-678 (2020).

[48] J. Hu and Y. Wang, The Clinical Characteristics and Risk Factors of Severe COVID-19, Geronotology, **67**, 255-266 (2021).

[49] T. Wu, P. Yu, Y. Li, J. Wang, Z. Li, J. Qiu, L. Cui, Y. Mou and Y. Sun, Asthma does not influence the severity of COVID-19: a meta-analysis, J Asthma, **59(6)**, 1188-1194 (2022).

[50] B. Huang, Z. Chen, M. Sidell, S. Eckel, M. Martin,ez, F. Lurmann, D. Thomas, F. Gilliland and A. Xiang, Asthma Disease Status, COPD, and COVID-19 Severity in a Large Multiethnic Population, Allergy Clin Immunol Pract, **9(10)**, 3621-3628 (2021).

Eman Khorsheed is an Associate Professor of Statistics and Data Science at University of Bahrain. She is a Fellow of the Higher Education Academy- UK. In 2021, she became an Elected Member of the International Statistical Institute (ISI). Most recently, she has been elected a Council Member of the International Society for Business and Industrial Statistics (ISBIS). She received the PhD degree at University of Bath (UK). She has published articles in reputed scientific journals. Her main research interests are: Bayesian statistics, and data science for modeling real-life problems in different sectors including Energy, Public Health, Environment, Industry, Economics, & Space Sciences. She has organized/participated in many international conferences.

Hawra Alshowaikh is a graduate of Cairo University, holding a Bachelor of Medicine, Bachelor of Surgery (MBBS) degree. Her primary research interests lie in the fields of medical science and data analysis.