Alexandria Engineering Journal 78 (2023) 162–174

Contents lists available at ScienceDirect

Alexandria Engineering Journal

journal homepage: www.elsevier.com/locate/aej

Original Article

Bayesian machine learning analysis with Markov Chain Monte Carlo techniques for assessing characteristics and risk factors of Covid-19 in Erbil City-Iraq 2020–2021

Hewir Abdulqadir Khidir [a], İlker Etikan [a,*], Dler Hussein Kadir [b,c], Nozad H. Mahmood [d], R. Sabetvand [e]

[a] Department of Biostatistics, Near East University, Nicosia, TRNC, Mersin 10, Turkey
[b] Department of Statistics and Informatics, Salahaddin University-Erbil, Kurdistan Region, Iraq
[c] Department of Business Administration, Cihan University-Erbil, Kurdistan Region, Iraq
[d] Department of Business Administration, Cihan University Sulaimaniya,Sulaimaniya, 46001, Kurdistan Region, Iraq
[e] Department of Mechanical Engineering, Khomeinishahr Branch, Islamic Azad University, Khomeinishahr, Iran

ARTICLE INFO

Keywords:
Covid-19
Bayesian neural network
Bayesian logistic regression modelling
Markov Chain Monte Carlo odds ratio
ROC curve
Classification measurements

ABSTRACT

The study aims to showcase machine learning techniques in the application of medical datasets for improving identification of correlations and relationships between variables, which will lead to more informed decision-making. Unlike other studies, intensive statistical modelling is used to understand and find the effective of variables cause to lead death due to Covid-19. Due to large dataset, not common approaches derive us to ideal conclusion. Furthermore, Bayesian technique is applied to generate predictive posterior distributions of the unknown parameters in the model in neural network as well as logistic regression, which helps us to avoid overfitting in machine learning applications and have additional measurements in assessing fitted model performance. According to the results extracted from the statistical analysis, the Bayesian neural network demonstrated superior performance in terms of classification measurements such as AUC (84.66%), F1 (87.11%), Precision (88.29%), and Recall (85.96%). The Bayesian logistic regression also performed well, but with slightly lower scores, achieving AUC (83.07%), F1 (85.59%), Precision (84.55%), and Recall (85.59%). In contrast, logistic regression (MLE) technique had the worst performance with very low scores (AUC = 52.38%, F1 = 57.55%, Precision = 57.01%, Recall = 58.10%). Regarding the variables' association with mortality, stepwise forward selection helped to identify seven significant variables. Age was found to be the most significant variable in predicting the probability of dying, with patients in the age group of (18–44) having 12 times higher odds, patients in the age group of (45–64) having 123 more odds, and patients above 65 years old having 436 times more chance to die compared to patients below 18 years old. Severe coughing was also significant with 7.26 odds, and patients suffering from diabetes had 2.82 times more chance to die. Moreover, SpO2 contributed to a decrease of 20% in the relative risk of dying from Covid-19 disease. Gender and Smoking did not show a significant association with mortality. Finally, the Bayesian approach showed higher sensitivity and specificity than the classic neural network.

1. Introduction

In December 2019, a newly emerged infectious disease called Coronavirus disease 2019 (Covid-19), which was caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was first identified in Wuhan, China. In Iraq, Covid-19 was first detected in Southern Iraq in February 2020 and as a result of the emergence of these incidents, the Kurdistan Region enacted stringent security measures. To determine the limit of spreading the virus in the region, several indicators were used such as, the closure of schools and colleges, closure of borders and airports, cancellation of civic and religious events, and obligatory quarantine for people returning from trips abroad and encounters. More than 1.2 million people have died because of Covid-19, a new coronavirus disease that has affected over 45 million individuals globally [1,2].

* Corresponding author.
E-mail address: iLker.etikan@neu.edu.tr (İ. Etikan).

https://doi.org/10.1016/j.aej.2023.07.052
Received 30 May 2023; Received in revised form 5 July 2023; Accepted 19 July 2023
Available online 25 July 2023

List of Notations

Terms	Definition
Covid-19	Corona Virus Disease 2019
ML:	Machine Learning
OR:	Odds Ratio
CI:	Confidence Interval
ROC:	Receiver Operating Characteristic
LR (MLE):	Classic Logistic Regression
LR (MCMC):	Bayesian Logistic Regression
BANN:	Bayesian Artificial Neural Network
ANN:	Artificial Neural Network
WHO:	World Health Organization
CR:	Creatine
WBC:	White Blood Cell
PLT:	Platelet
SpO2	Oxygen
DIC:	Disseminated Intravascular Coagulation
FDPs	Fibrin Degradation Products
CAD	Coronary Artery Disease
TMPRSS2	Transmembrane Serine Protease 2
ACE2	Angiotensin-Converting Enzyme 2
MCMC	Markov Chain Monte Carlo
MH	Metropolis-Hastings
HMC	Hamiltonian Monte Carlo
KL-divergence	Kullback-Leibler Divergence
ELBO	Evidence Lower Bound
SVI	Stochastic Variational Inference
AUC	Area Under ROC Curve
ROC	Receiver Operating Characteristic Curve

A broad spectrum of symptoms can be caused by Covid-19, as 14 percent and 5 percent of the patients whose test results were confirmed had Covid-19 cases that were either severe or serious. The emerge damaged the healthcare system because it spread very quickly and was the main cause of death, consequently most of the available medical resources were secured to contain the virus.

Some of the common symptoms identified initially were dyspnea, breathing rates below 30 breaths per minute, blood oxygen saturation levels under 93%, partial pressure of arterial oxygen to fraction of inspired oxygen ratio of$<$300 mmHg, and/or lung infiltrates of over 50%. These indicators typically appeared within 24 to 48 h. Patients infected with severe respiratory issues required mechanical assistance for breathing and needed to be transferred to the critical care unit due to conditions such as shock, disseminated coagulopathy, or multiple organ failures.

Numerous risk factors were considered to be effective including age, gender, ethnicity, as well as nutrition, lifestyle, and laboratory indicators. It was generally agreed that the risk factors could help identify individuals who were more likely to contract the virus severely and face a higher risk of mortality. However, it was crucial to acknowledge that certain studies examine general risk factors associated with disease progression, whereas others focus on specific risk factors contributing to the advancement of Covid-19 into a critical stage [3]. Therefore, accurate predictions had became point of concern to contain the disease and find proper medicine for patients. Bayesian statistical methods plays crucial roles and alongside of classical approach for model fitting, and several studies have been studied to investigate Covid-19 risk factors as well as the trend of uprising the spread by [4–8]. This is because rather than providing a single-point estimate for the unknown parameters, Bayesian methods also offer the whole posterior distributions for the unknown parameters and Markov chain Monte Carlo sampling is one of the prevalent approaches to calibrating Bayesian models. Also, artificial intelligence and machine learning techniques were compared with artificial neural network as studied by [9–14]. Moreover, Bayesian calibration is a more accurate method, and due to the complexity of the likelihood function, it is difficult to calculate the model's parameters straightforwardly. Accurate Bayesian computation (ABC) can be applied to estimate the posterior distribution of unknown parameters when the likelihood function is hard to calculate or not known. ABC approaches are known for being inaccurate at dealing with large amounts of data and complicated models. This means that they can only be used for very simple models with a limited number of parameters. In addition, classic neural network may lead to overfitting when large number of inputs are entered in the model. In this study, we developed Bayesian neural network by involving Gibbs sampler while choosing prior values as well as hyperparameters which provided more accurate and fast convergence in the later stages of the HMC algorithm. Moreover, the contribution of the significant variables has not been pointed out in the current existed literature, thus all effective factors are stated in this study.

When individuals contract SARS-CoV-2 and become severely ill, the pulmonary system is often the primary organ which is impacted. Nevertheless, the virus has the potential to impact any organ in the body and can result in effecting multiple organs. It's important to assess how the organs are being impacted when treating patients. Disturbances in coagulation and vascular endothelium can cause injury to multiple organs, even though no symptoms appears during the early stages. Fig. 1 illustrates the parts of the human body that are affected by Covid-19.

Covid-19 can cause an excessive release of cytokines, which can result in systemic inflammation, multi-organ injury, and even death [15]. This response is known as cytokine storm or hypercytokinemia. SARS-CoV-2 has the potential to distort endothelial cells in different parts of human's body, resulting in inflammation, edema, vasoconstriction and hypercoagulability. These changes contribute to organ ischemia. [16]. Referencing Wang, Nie [17], the inflammatory response may persist even as the viral load decreases. The risk of clotting is increased by blood vessel constriction, hypercoagulability, immobilization, endothelial, inflammation and hypoxia [18]. Zhang, Xiao [19] pointed out it also causes moderate thrombocytopenia and elevated C-reactive protein. Furthermore, they also highlighted the effects of the virus on other body parts like DIC, lymphocytopenia, fibrin degradation products (FDPs) and D-dimer, and Covid-19 prognosis can be predicted by D-dimer levels and DIC.

Skin symptoms in Covid-19 resemble those of other viral infections and inflammatory of chronic conditions, for example, psoriasis, acne, rosacea, and eczema. Skin manifestations accompanied by vascular complications may be attributed to immune complex causes, microthrombotic, or neurogenic. A majority of Covid-19 infected cases with skin symptoms present with patchy erythematous rash, while others exhibit hives or widespread urticaria. The presence of skin manifestations does not indicate a more severe form of Covid-19. Cardiac complications in Covid-19 can occur independently from pulmonary and other issues [20,21]. Individuals with preexisting coronary artery disease (CAD), latent CAD, or no CAD are at risk for ischemic cardiac injury caused by plaque rupture and thrombosis or inadequate oxygen supply.

Autopsy studies have shown that individuals experiencing the acute stage of Covid-19 demonstrate a standard prolix alveolar injury that lacks organization and fibrosis. This type of damage occurs due to the disruption of endothelial and alveolar cells, which leads to the exudation of fluid and cells and the formation of hyaline membranes. [22]. Colavita, Lapa [23] derived that cells in the ocular surface of the eye, including those in the cornea, within the conjunctiva, and in the sclera of the eye, possess both TMPRSS2 proteases and ACE2 receptors, which are necessary for SARS-CoV-2 infection. Roughly one-third of hospitalized individuals encounter ocular abnormalities, such as conjunctivitis, and

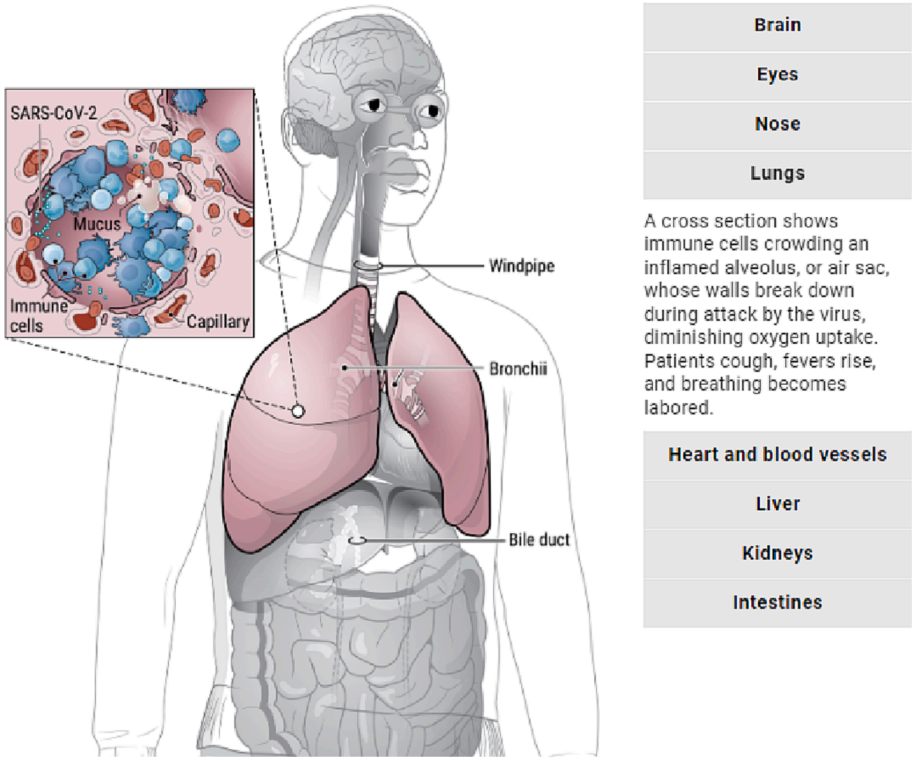

Fig. 1. Human's Body with description of Covid19 Impact.

this occurrence is more prevalent in individuals with more severe illnesses. Ocular involvement can manifest in the initial stages of the illness, and cells on the ocular surface are potential entry points and reservoirs for the virus. Shedding of the virus in ocular secretions can be a source of transmission, and the virus can sustain its infectivity in the eye for a duration of up to three weeks.

Likewise, Covid-19 can cause a suppression of brain stem reflexes, including the reflex responsible for detecting oxygen deprivation. Neurological symptoms may either be the sole presenting symptoms or present in combination with respiratory or other symptoms and are more frequent in severe cases of the disease. Abnormal levels of oxygen and carbon dioxide may contribute to these symptoms, and this could be delirium, headaches, confusion, dizziness, altered consciousness, and difficulty awakening [24].

The article is structured as follows: In Section 2, the concept of statistical Bayesian learning is highlighted, shedding light on their significance and applications. This section provides a comprehensive understanding of the theoretical foundations. To validate the chosen approach, Section 3 presents a practical application, offering a comparative analysis to showcase its effectiveness. The application of the approach to a real-world problem adds credibility to the research. Section 4 serves as the discussion where intensive arguments are provided, summarizing the key findings and implications of the study. Finally, the rest is for limitation, conclusion as well as recommendation which offers insights into the broader implications and potential future directions.

2. Methodology and statistical learning of Bayesian inference

2.1. Methodology

This chapter presents the patient's profile and methodology, including medical history, laboratory results and the person's demographics. No personal information about the patients is explored such as names, phone numbers, or addresses. The process of discovering patterns, correlations, changes, deviations, and statistically significant structures and events in large datasets is considered. Distributions that characterize an observable property (descriptive statistics) are generated by classical statistical methods and used to assess the validity of a sample drawn from a larger population. To optimize Neural Network calculations, we present a novel metric that incorporates MCMC approaches and then evaluate the results against the standard approach.

2.2. Research method

2.2.1. Site of study

We have used secondary data of a cross-sectional study type since the information was already collected by the hospitals themselves. The data includes patients from 2020 to 2021. SPSS Version 25 software, Python as well as R code-based programming language and writing materials have been used to generate the results.

2.2.2. Study population

The target population comprised patients of Covid-19 taken from the hospital in the recorded of 2020 to 2021. The study population comprised the Bayesian ANN analysis with MCMC approaches of Assessing Characteristics and Risk Factors of Covid-19 cases.

2.3. Data capture and analysis strategies

2.3.1. Data cleaning

The analysis is primarily impacted by many missing participants in our data, and we, thus, have removed most extreme and unpredictable values from the dataset. In addition, the data has also been verified for consistency. The data from patients' files are captured in a Microsoft Excel spreadsheet which is $n = 537$.

2.3.2. Data analysis

Once the data set was cleaned and organized, we utilized R software to import and analyze the data. The frequencies and percentages of all

variables were calculated using R software. As we already defined that our dependent variable Covid-19 risk factors are in binary form as discharged alive/died, the logistic regression model was first applied. Due to existing the association of these different risk factors, statistical approaches that analyse their relationships were needed to examine their effects as well as multicollinearity. Our new proposed term is to involve MCMC techniques for optimization in Neural Network calculation and then compare it to classic neural network with single point estimation.

2.4. Bayesian logistic regression model

In recent years, the Bayesian inference framework has gained popularity as a more appealing method for estimating parameters in logistic regression. This approach offers easier interpretability of parameter estimation and yields more dependable results for smaller samples. Bayesian inference provides a valuable approach for merging expert knowledge, also known as prior beliefs with data to generate posterior distribution. Consequently, in the event of new data being gathered, the Bayesian framework can be utilized to revise current knowledge by incorporating fresh data. This updating process can be repeated as additional data is accumulated in the future. The fundamental principle behind all Bayesian inference is Bayes' theorem, which has been extensively explored in literature by scholars such as[25–29]. To comprehend the utilization of the Bayesian approach in parameter estimation, let β denote the vector containing k unknown parameters and let X represent the vector containing n observations.

$$\beta = (\beta_1, \beta_2, \cdots, \beta_k)$$

$$X = (x_1, x_2, \cdots, x_n)$$

Based on Bayes' theorem, the posterior probability distribution $P(\beta/X)$ can be expressed.

$$P(\beta/X) = \frac{p(\beta)^* p(X/\beta)}{\int p(\beta)^* p(X/\beta) dX} \tag{1}$$

The likelihood and Log Likelihood Function.

The logistic regression model's likelihood is determined as follows:

$$f\left(\frac{y}{X}, \beta\right) = \prod_{i=1}^{n} p^{y_i} (1-p)^{1-y_i} \tag{2}$$

$$logf(y/X, \beta) = \sum_{i=1}^{n} -y_i(p) + (1-y_i) log(1-p) \tag{3}$$

By replacing p with $\left[\frac{1}{1+e^{-X\beta}}\right]$ and doing some calculations, the above formal becomes

$$logf(y/X, \beta) = \beta X(y - 1_n) - 1_n \left[\log\left(1 + 1 + e^{-X\beta}\right)\right] \tag{4}$$

We employ a normal distribution with a mean of zero for each unknown parameter as a starter, which leads to the prior distribution taking the shape of a normal distribution. We propose a multivariate normal prior for β:

$$\beta\ N\left(0, \sigma_{\beta}^2\right)$$

Thus, its probability density function log without constant terms is as followings:

$$logp\left(\beta \middle/ \sigma_{\beta}^2\right) = -\frac{1}{2} log\sigma_{\beta}^2 - \frac{\beta^T \beta}{2\sigma_{\beta}^2} \tag{5}$$

The log of the posterior distribution is driven below:

$$logf\left(\beta \middle/ X, y, \sigma_{\beta}^2\right) \propto \beta X(y - 1_n) - 1_n \left[\log\left(1 + e^{-X\beta}\right)\right] - \frac{\beta^T \beta}{2\sigma_{\beta}^2} \tag{6}$$

As a result, the gradient function of the leapfrog function can be written as:

$$\Delta_{\beta} \log f\left(\beta, X, y, \sigma_{\beta}^2\right) \propto X\left(y - 1_n + \frac{e^{-X\beta}}{1 + e^{-X\beta}}\right) - \frac{\beta}{\sigma_{\beta}^2} \tag{7}$$

2.4.1. Markov Chain Monte Carlo: The basics

Markov Chain Monte Carlo (MCMC) is a diverse category of computational techniques for estimating integrals and generating posterior samples. In Bayesian analysis, MCMC algorithms are mainly employed to approximate the posterior distribution by generating simulated samples. The Metropolis-Hastings (MH) algorithm is a widely used general principle for generating posterior samples in Bayesian analysis. The Gibbs technique sampler is a specific instance of the MH process [30].

2.4.2. Hamiltonian Monte Carlo algorithm

By utilizing a guided proposal generation scheme, Hamiltonian Monte Carlo (HMC) enhances the efficiency of the MH algorithm. HMC achieves this by leveraging the gradient of the logarithm of the posterior distribution, which guides the Markov chain towards areas of higher posterior density where most samples are concentrated. Therefore, a well-optimized HMC chain is capable of accepting proposals at a significantly higher rate compared to the traditional MH algorithm [31]. Further detailed explanations can be found in other sources, such as [32,33]. The Hamiltonian function, denoted as $H(\beta, p)$, is expressed as the sum of potential energy $U(\beta)$ and kinetic energy $K(p)$, where β and p are both in the real $k-dimensional$ space, i.e., $\beta, p \in \mathbb{R}^k$. Specifically, the expression is given as $H(\beta, p) = U(\beta) + K(p)$. With this formulation, we possess/obtain.

$$H(\beta, p) = -logf(\beta) + \frac{1}{2} p^T M^{-1} p. \tag{8}$$

The full package of the Hamiltonian technique is provided in the Algorithm below:

Hamiltonian Monte Carlo Algorithm

$Input(\beta^{(0)}, logf\left(\beta/X, y, \sigma_{\beta}^2\right), M, N, L, \epsilon)$
$Settingstartingpointforlogf\left(\beta/X, y, \sigma_{\beta}^2\right)$
$fort = 1 toN$
$p = Rand.noramal(0, M)$
$\beta^{(t)} = \beta^{(t-1)}, \widetilde{\beta} = \beta^{(t-1)}, \widetilde{p} = p$
$fori = 1 toL$
$\widetilde{\beta}, \widetilde{p} \rightarrow Leapfrog\left(\widetilde{\beta}, \widetilde{p}, \epsilon, M\right)$
$Endfor$
$\propto = min\left(1, \frac{\exp\left(\log f\left(\widetilde{\beta}\right) - \frac{1}{2}\widetilde{p}^T M^{-1} \widetilde{p}\right)}{\exp\left(\log f\left(\widetilde{\beta}^{(t-1)}\right) - \frac{1}{2} p^T M^{-1} p\right)}\right)$
$withprobability \propto, \beta^{(t)} = \widetilde{\beta} and p^{(t)} = -\widetilde{p}$
$Endfor$
$Return \beta^{(1)}, \beta^{(2)}, \cdots \beta^{(N)}$
Run (LeapFrog Function)

2.5. Bayesian neural network approach

A Bayesian neural network (BNN) refers to a type of neural network architecture that integrates Bayesian inference principles. Similar to traditional neural networks, a BNN consists of layers of interconnected nodes, or neurons, that receive inputs, apply weights and biases, and produce outputs through activation function. However, unlike traditional neural networks, a BNN is designed to model not only the mapping from inputs to outputs but also the uncertainty associated with that mapping. Fig. 2 demonstrates the architecture of both Bayesian neural networks as well as neural network with point estimation. In a BNN, the biases and weights of the neural network are treated as random variables with prior distributions. The prior distributions reflect the prior belief or uncertainty about the values of the weights and biases. Bayesian

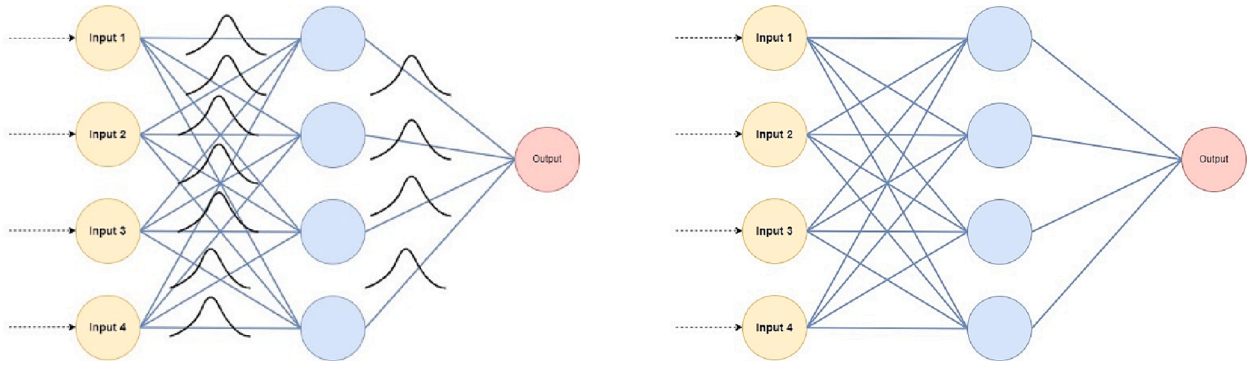

Fig. 2. Left: Bayesian Neural Network with probability distribution over weights. Right: Classic Neural Network with point estimates for weights.

inference aims to update the prior probability distributions using the observed data, resulting in posterior distributions that reflect the updated beliefs or uncertainty about the values of the weights and biases.

2.5.1. Bayesian learning process

The Bayesian learning process begins with the definition of a model, M, and a prior distribution $p(w)$ for the model parameters. After examining additional data used to update, then the updated distribution of priors is utilized to create the posterior distribution using Bayes' rule.

$$P(W/X) \propto p(W)*p(X/W) \tag{9}$$

2.5.2. Likelihood

In statistics, the likelihood function is driven from its original probability density function for n data points in multiplying the function for each joint (x,y) dataset. Its general written formula is as followings:

$$L(W/X) = \prod_{i=1}^{n} p\left(y^{(i)}|X^{(i)}, W\right) \tag{10}$$

To write in more detail explicitly for two classes as in our cases where the sigmoid function is intended to be used for the output layer, Equation (11) provides more insights on the likelihood function:

$$p(y=1|X,W) = [1+\exp(-f(x,w))]^{-1} \tag{11}$$

Free parameters which are called (weights) in the model must be set corresponding to the training set size, the noise level as well as the target function complexity when computing classical estimation (error minimization) for the Multi-Layer Perceptron (MLP). However, limiting the network size is no longer an issue in the Bayesian technique, but it is wise to minimize the number of hidden units in practice for computational purposes. In addition, referring to Neal [32], for small sample sizes, the process of converging tends to be Gaussian while implying limiting hidden unit numbers which is considered as a feasible practice in such circumstances. The predictive distribution is achieved by calculating the model's integration of predictions regarding the posterior distribution to predict the new output $y^{(n+1)}$ when new values of input $x^{(n+1)}$ is available.

$$\widehat{y}^{(n+1)}{}_{k} = \int f_k\left(X^{(n+1)}, w\right) p(w, \alpha, \tau/X) dw \alpha \tau \tag{12}$$

2.5.3. Hybrid Monte Carlo algorithm for BNN

The Hybrid Monte Carlo (HMC) algorithm is a computational method used in statistical physics, Bayesian statistics, and other fields to sample from complex probability distributions. The algorithm combines molecular dynamics simulations with Markov Chain Monte Carlo (MCMC) sampling to explore high-dimensional spaces efficiently. Equation (12) represents the expectation of function $f(X^{(n+1)}; W)$ with regards to the parameter's posterior distribution. The Monte Carlo method can implement to approximate this by drawing a sample of values $W^{(t)}$ from the posterior distribution.

$$\widehat{y}_k^{(n+1)} \approx \frac{1}{N} \sum_{t=1}^{N} f_k\left(X^{(n+1)}, w^{(t)}\right) \tag{13}$$

In the analysis process, the hybrid Monte Carlo (HMC) algorithm is applied to compute the parameters, while Gibbs sampling is used to calculate the hyperparameters. HMC is a powerful Monte Carlo method that leverages gradient information to reduce random walk behavior commonly observed in the Metropolis algorithm. The gradient provides direction for jumping to states with higher probability. Additionally, Gibbs sampling is used to estimate hyperparameters, which helps to reduce the amount of tuning required for HMC to achieve satisfactory performance.

2.5.4. Variational inference

Variational inference aims to approximate the exact posterior distribution p(W/X) using a parameterized distribution $q_\theta(W)$, called the variational distribution, instead of sampling from it directly. This variational distribution is defined by a set of weight parameters that are learned to make $q_\theta(W)$ as similar as possible to $p(W/X)$. To measure this similarity, the Kullback-Leibler divergence (KL-divergence) is commonly used [34], which is based on Shannon's information theory [35] and measures the distance between two probability distributions.

In Bayesian inference, the KL-divergence measures the number of additional bits, on average, that are needed to encode a sample from the true posterior distribution using a code optimized for the approximate posterior distribution. This can be expressed mathematically as:

$$D_{KL}(q_\theta(W)\|p(W/X) = \int_{\varnothing} q_\theta(W) log \frac{q_\theta(W)}{p(W/X)} dW \tag{14}$$

To circumvent this, a distinct formula known as the evidence lower bound (*ELBO*) can be employed as a loss function, which is straightforward to derive.

$$\int_{\varnothing} q_\theta(W) log\left(\frac{p(W,X)}{q_\theta(W)}\right) dW = \log(p(X) - D_{KL}(q_\theta(W)\|p(W/X) \tag{15}$$

Since $log(P(X))$ is exclusively determined by the prior, the minimization of $D_{KL}(q_\theta(W)\|p(W/X)$ is equal to the maximization of the evidence lower bound (*ELBO*).

2.5.5. Bayes by backpropagation

Variational inference is a powerful tool for the Bayesian inference method, but it needs to be adapted to work in the deep learning context. The primary challenge is that stochasticity prevents backpropagation from working properly at the internal nodes of a network [36]. To address this issue, several solutions have been proposed, such as probabilistic backpropagation [37] and Bayes-by-backdrop [38]. Bayes-by-backprop is a practical implementation of stochastic variational

inference (SVI) combined with a reparameterization trick [39], which ensures that backpropagation utilizes ordinarily. The full package of the Bayes-by-backprop technique is provided in the Algorithm below:

Bayes by Backprop Algorithm

$Set\varnothing = \varnothing_0$
$\boldsymbol{for\ t} = 1\boldsymbol{to}N$
$sample\epsilon\ q(\epsilon)$
$W = t(\epsilon, \varnothing)$
$f(W, \varnothing) = \log(q_{\varnothing}(\mathrm{W})) - \log(p(y/X, W)P(W)$
$\Delta_{\varnothing} f = backprop_{\varnothing}(f)$
$\varnothing = \varnothing - \alpha\Delta_{\varnothing} f$
$\boldsymbol{Endfor}$

3. Statistical analysis results

3.1. Descriptive and visualization analysis

It is begun with descriptive statistics since it is considered a vital stage of any study's outcomes where initial thoughts on the nature of the dataset are found and sometimes enabling the researchers to find potential patterns among the explanatory and response variable. To simplify the discussion and easy to follow, it was split into two parts, exploring the association of predicted variables with quantitative and qualitative covariates separately.

3.1.1. Response association with quantitative variable

Table 1 provides attention to the laboratory measurements which affects on the response variable and whether the increase or decrease unit of any of them caused Covid-19 patients to die. For instance, the mean value of HR for those discharged alive in the hospital was measured at (84.7821 ± 13.7939) while it was (107.9230 ± 16.4631) for those who died, thus it led us to report that the disease had an impact on increasing heart rate pulse. Additionally, a notable disparity in mean values was observed for the CRP variable, with 16.6035 for recovered cases and 48.3684 for deceased cases. Furthermore, when compared to upper respiratory tract infections, individuals who succumbed to Covid-19 exhibited a significantly elevated mean value. (22.6862 ± 4.6901 vs 18.8161 ± 1.8521, p-value < 0.001), and mean value of quadrant (2.5161 ± 1.0671 vs. 3.1837 ± 1.1263, p-value < 0.001) and pulmonary (33.2757 ± 17.8723 vs. 39.7602 ± 16.9315, P-value < 0.001) were slightly higher and significantly differed. As stated, no significant difference occurred due to Neutrophil results between died and recovered cases with (3.1601 ± 0.4974 vs. 3.1403 ± 0.6824, p-value = 0.6993). Upon admission, individuals who eventually died from severe/critically ill Covid-19 exhibited higher body temperatures, lower SpO2 levels, as well as pulmonary opacity values and higher CT image quadrant scores.

In addition, infected cases who died from Covid-19 had higher fibrinogen, C-reactive protein and D-dimer levels than recovered cases. Similarly, those who passed away had lower APTHT, lymphocyte, platelet, and albumin count as per the results provided in Table 1. Fig. 3 illustrates the distribution of the variables against the response variable where one can simply identify the effectiveness of the covariates on the dependent variable. For example, regarding CRP, HR, SpO2, Lymphocyte and WBC_Count measurements, the distribution of the survived and dead cases was split for two different areas with a very low rate of overlaps, and these variables were already providing insights to be listed on predicting the probability not surviving a patient. However, it was not wise to decide at this stage to highlight variables that had an impact on increasing the risk of dying from the disease since further tests are required to be implemented.

Table 1
Descriptive statistics of laboratory results under investigation recovered and died cases.

Variables	Recovered		Died		P-Value[a]
	N	Mean ± SD	N	Mean ± SD	
Temperature	341	37.5560 ± 1.9746	196	38.2954 ± 1.6640	0.0000
HR	341	84.7821 ± 13.7939	196	107.9230 ± 16.4631	0.0000
Respiratory	341	18.8161 ± 1.8521	196	22.6862 ± 4.6901	0.0000
Quadrant	341	2.5161 ± 1.0671	196	3.1837 ± 1.1263	0.0000
Pulmonary	341	33.2757 ± 17.8723	196	39.7602 ± 16.9315	0.0000
Neutrophil	341	3.1601 ± 0.4974	196	3.1403 ± 0.6824	0.6993
Lymphocyte	341	1.3683 ± 0.2730	196	0.7750 ± 0.2106	0.0000
Platelet	341	192.9123 ± 22.9700	196	172.7015 ± 27.0225	0.0000
Albumin	341	41.4877 ± 2.2717	196	38.5495 ± 2.2308	0.0000
Creatinine	341	66.6337 ± 7.4247	196	68.2755 ± 9.1109	0.0238
APTHT	341	33.5270 ± 2.5521	196	32.2791 ± 2.2824	0.0000
Fibrinogen	341	3.3686 ± 0.3970	196	4.4469 ± 0.8309	0.0000
SpO2	341	96.6765 ± 3.6453	196	91.1806 ± 4.4007	0.0000
WBC_Count	341	4.9604 ± 0.9252	196	3.8378 ± 0.7627	0.0000
CRP	341	16.6035 ± 19.9936	196	48.3684 ± 30.8121	0.0000
D_dimer	341	0.4493 ± 0.2345	196	0.6235 ± 0.2383	0.0000

[a] Continuous variable: T-test or Mann-Whitney tests as appropriate.

3.1.2. Response association with qualitative variables

The study calculated figures for nine categorical variables were counted to estimate the likelihood of mortality due to Covid-19. Table 2 calculates the ratio of death rates per gender (male death rate: female death rate) in Covid-19 patients. The male death rate was 25% times higher than the overall female death rate. This means that according to the cohort included in this study, men make up to 20.5% of all Covid-19 deaths while only 16% of the death rate was recorded as women.

Concerning Smoking factor, a total of 537 Covid-19 infected individuals are included and observed in this study, 196 of whom (36.5%) experienced disease progression and 214 (40%) with a history of smoking. Among those with a history of smoking, 13.4% experienced disease progression and died, compared with 23.1% of non-smokers. The analysis revealed no significant association between ever smoking and Covid-19 progression. Furthermore, the statistical results indicated that stroke was not significantly linked to Covid-19 mortality, as displayed in Table 2. The mortality rate among patients who had a stroke was lower than those without stroke which is 13.8% and 22.7% respectively. On the contrary, patients with no previous history of stroke had shown greater improvement and recovery which is 38.6% compared to recovery among patients who had previous history of stroke which is 25%.

3.1.3. Model building Analysis

To avoid multicollinearity issues which result high p-value as well as unreliable estimated parameters while building up models with especially multiple variables which of course influence the predicted values afterward. It was most likely believed that there might be a high correlation among the covariates and referencing Fig. 4, it was noted that Fibrinogen had a moderate and positive relationship with Age and.

3.2. Model fitting

Forward selection was used to determine which variables should be included in the final model, starting with a simple null assumption. First, the null model with (Status ~ 1) was calculated, and its residual deviance was recorded. This model has (n-1) degrees of freedom, where n refers to the total number of observations. This model was assumed to be poor, so additional analysis was required. Then, for each explanatory variable, only the response variable was used, yielding Status ~ Age,

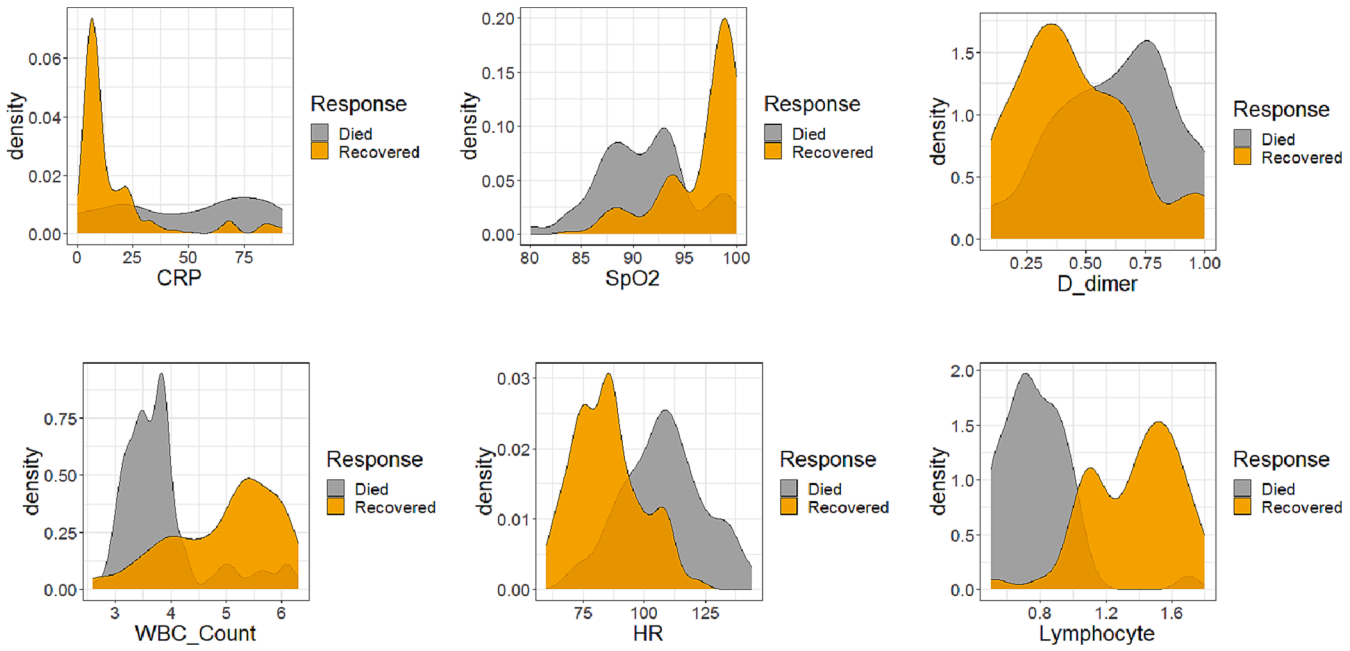

Fig. 3. Density Distribution for the response variable.

Table 2
Descriptive Statistics of Categorical Variables Associated with Response Variable.

Variables	Levels	Recovered		Died		P-Values[b]
		N	%	N	%	
Gender	Male	199	37.1%	110	20.5%	0.6140
	Female	142	26.4%	186	16.0%	
Age	<18	98	18.2%	4	0.7%	0.0000
	19–44	167	31.1%	39	7.3%	
	45–64	53	9.9%	55	10.2%	
	65+	23	4.3%	98	18.2%	
Smoking	Yes	142	26.4%	72	13.4%	0.2630
	No	199	37.1%	124	23.1%	
Fever	Yes	137	25.5%	112	20.9%	0.0000
	No	204	38.0%	84	15.6%	
Cough	Yes	142	26.4%	126	23.5%	0.0000
	No	199	37.1%	70	13.0%	
Sputum	Yes	99	18.4%	75	14.0%	0.0280
	No	242	45.1%	121	22.5%	
Hypertension	Yes	135	25.1%	133	24.8%	0.0000
	No	206	38.4%	63	11.7%	
Diabetes	Yes	119	22.2%	133	24.8%	0.0000
	No	222	41.3%	63	11.7%	
Stroke	Yes	134	25.0%	74	13.8%	0.7240
	No	207	38.5%	122	22.7%	

[a]Categorical variables: Fisher Exact or Chi-square tests as appropriate.

Status ~ Gender, and so on. Each model was evaluated by calculating its residual deviance, and the one with the lowest value was selected for further analysis. The model with the Age variable included was chosen to the next phase and proceeded the same as above where their AIC as well as Nagelkerke R-Sqaure were measured and recorded as shown in Table 3. As a result, this analysis led to the identification of the 7 most significant variables.

Table 4 illustrates the best-fitted model from both approaches. Similar to univariate outputs, the same conclusion can be made where the Bayesian approach performed better according to the standard errors of the coefficients, although at Age (2) and Age (3) the MLE technique produced lower SE of the parameters.

3.2.1. *Parameter interpretation*

Tables 4 and 5 are the most important part of logistic regression modeling where the magnitude of the variables can be identified. To begin with, Age (18–34) coefficient (2.2171) which stands for 18–35 years old provided in Table 4, is statistically significant (associated with a p-value of 0.05), implying that the Age factor does influence the risk of being died from Covid-19 disease. Because it is a positive number, we can conclude that age raises the risk of developing the disease. Therefore, the odds ratio of Age (18–34) was calculated as approximately 12 with 95%CI (0–0) as shown in Table 6. This means holding other variables as constant, a patient in the age group (18–44) had 12 times higher chance of losing their life because of Covid-19 compared to individuals who were<18 years old, and people in the age group (45–64) had 123 more odds to die as well as 436 times more chance in age group more 65 years old than patients were<18 years old. It can be noticed that younger people had a higher chance to survive the disease. Similarly, coughing symptoms severely identified among admitted patients had a significant impact on increasing the odds to die. This pointed to the that inpatient with coughing will rise the odds of dying by Exp (1.9824) = 7.26 times as seen in Table 5. That being said, an inpatient with coughing had 7 times higher chances to die compared to those not having a strong cough. Like coughing, diabetes was also found to be statistically significant with a coefficient (1.2976) and p-value < 0.001, and recognized to increase the risk's odds by Exp (1.2976) = 2.8287. This enables us to report that there was a 182% increase in the odds of passing away with the presence of diabetes. In addition, people possessing Covid-19 as well as hypertension had high risk and pointing to the result, the factor had a positive coefficient value with (2.0048) and was statistically significant under 0.05 level which led to an increase in the logit of predicting dying. Thus, the odds of a patient died who had hypertension was 6.6 times higher than patients who did not suffer from hypertension with a 95% CI of (2.495 and 12.097). Moving to SpO2 which refers to measured oxygen for hospitalized cases, with a coefficient (-0.2132), and its odds ratio was (0.80). Hence, SpO2 is associated with a 20% (1 – 0.80 = 0.20) reduction in the relative risk of dying. In addition to that, for a 1-unit increase in the corresponding oxygen's level of patient admitted to hospital due to the disease is associated with a lower risk of dying due to Covid-19.

Referencing the effect of white blood cell count on the probability of dying holding other variables as constant, the estimated parameter was

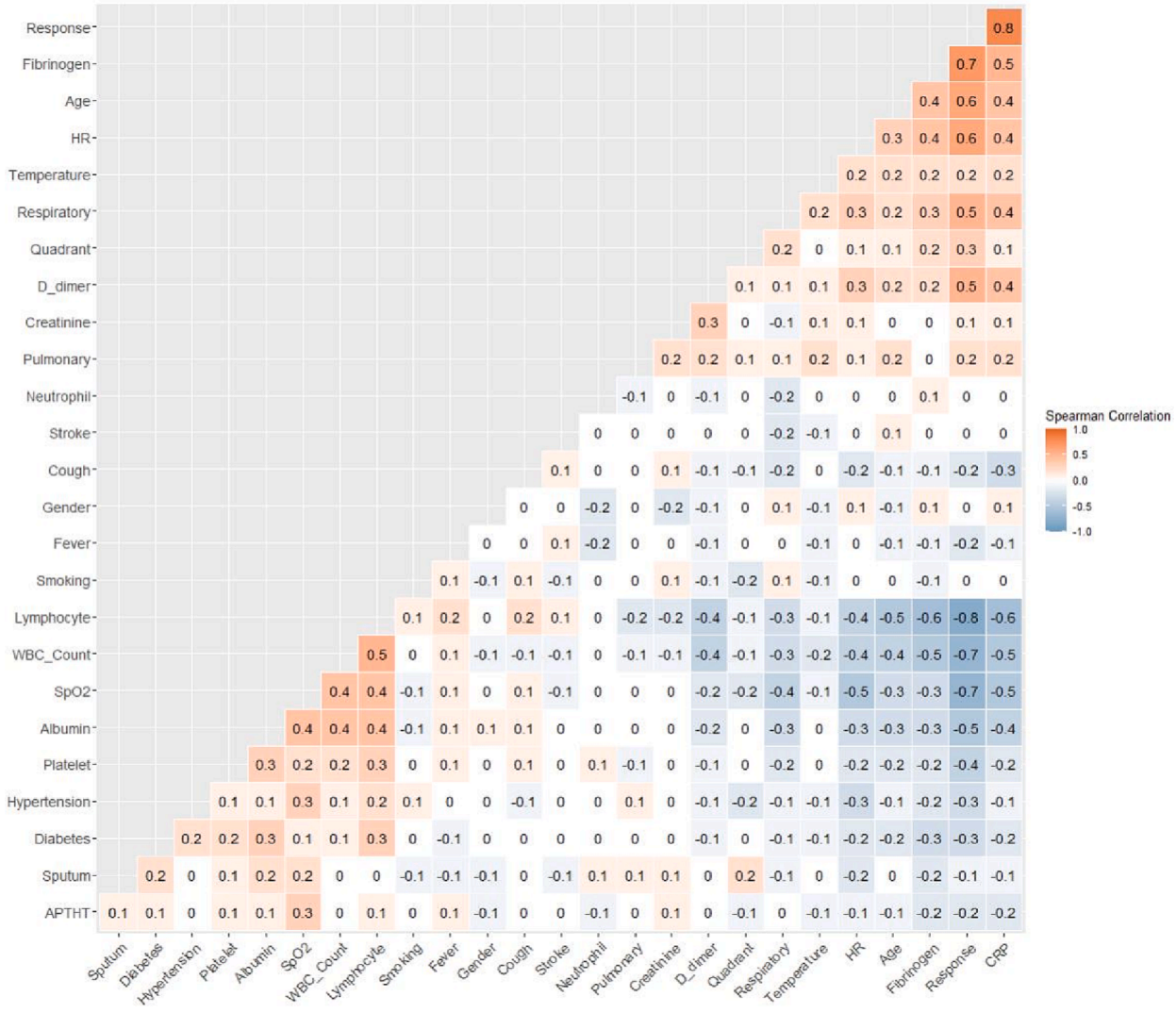

Fig. 4. Correlation Heatmap Matrix for all variables study.

Table 3
Result of stepwise forward model selection approach for Bayesian logistic regression (MCMC) and classic logistic regression (MLE).

Models	MLE Approach		MCMC Approach	
	AIC	Nagelkerke R-Sqaure	AIC	Nagelkerke R-Sqaure
Model 1	314.500	48.000%	308.4967	47.970%
Model 2	224.700	68.700%	216.2335	68.660%
Model 3	183.970	76.500%	174.1632	76.430%
Model 4	171.490	78.900%	159.7710	78.870%
Model 5	160.870	81.000%	148.3420	82.310%
Model 6	146.35	83.600%	132.4520	84.643%
Model 7	142.21	84.500%	129.3280	85.212%

(-0.9886), which means that with one unit increase from white blood cells, the logit of predicting dying decreases. Turning to the odds ratio, it was estimated to be 0.37; for every 1 increase in white blood cell count, an infected Covid-19 person had a 63% chance to survive than not with a 95% CI of (0.195 and 0.551). The CRP turned out to be contributing in increasing the odds of dying with a coefficient (0.0225), which is positive. One can infer that an increase in CRP was linked to a higher probability of death from Covid-19. It was important to state that CRP was associated with a 2.72% rising the chance to die. Although the effect size was small, it was statistically significant with a p-value < 0.001.

3.2.2. *Traceplots and histograms illustration of posterior distribution*

Various methods exist to calculate Bayesian credible intervals, but for this study, the 95% intervals were computed for each of the parameters. Table 6 displays the 0.025 and 0.975 quantiles for each parameter, as well as the 25th, 50th, and 75th percentiles. The credible intervals for each parameter can be determined by listing the values in the first and last rows. In particular, the credible interval for Age is positively skewed and does not encompass zero, implying that there is evidence that the Age group was positively related to the odds of not surviving the disease. However, SpO2 as well as WBC were negatively associated with the response variable confirming a negative link. The credible intervals also provide us insights into the posterior distribution and whether can be reliable or not.

To start, examine the traceplots of the first 100 trials after the burn-in period (refer to Fig. 5). In addition, a histogram can be used to represent the posterior distribution of each unknown parameter, which gives an idea of the parameter's marginal distribution.

There are additional ways to summarize our findings such as minimum, maximum, mean, and standard deviation. Table 7 shows some basic statistical measures of the posterior distribution and enables us to understand the distribution easier.

3.2.3. *Model accuracy and diagnosis assessment*

Low values indicate that the model is a better fit to the data in both

Table 4
The best-fitted model outcomes by Bayesian logistic regression (MCMC parameter estimation).

Included Variables	Classic Logistic Regression (MLE)			Bayesian Logistic Regression (MCMC)		
	Coefficients	SE	P-value	Coefficients	SE	P-value
Intercept	28.6433	6.1532	0.0000	27.6297	4.2894	0.0000
Age (18–34)	2.2171	0.9668	0.0218	2.4636	1.0315	0.0000
Age (35–65)	4.8459	1.0393	0.0000	4.8139	1.0792	0.0000
Age (>65)	6.0925	1.0975	0.0000	6.0791	1.0708	0.0000
SpO2	−0.3333	0.0655	0.0000	−0.2132	0.0453	0.0000
WBC_Count	−1.1157	0.2649	0.0000	−0.9886	0.2441	0.0000
Diabetes (Yes)	1.2976	0.5109	0.0111	1.0398	0.4518	0.0000
Cough (yes)	2.0109	0.5858	0.0006	1.9824	0.5305	0.0000
Hypertension (Yes)	2.0048	0.5565	0.0003	1.8886	0.4864	0.0000
CRP	0.0225	0.0093	0.0156	0.0268	0.0085	0.0000

Table 5
Odds Ration and 95% Confidence Interval of Odds Ratio for Bayesian logistic regression coefficients estimated by MCMC.

Variables in the model	Odds Ratio (MCMC Logistic Regression)	95% CI of Odds Ratio	
Age (18–44)	11.7469	9.380	17.069
Age (45–64)	123.2075	106.592	175.452
Age (>65)	436.6317	421.492	521.402
SpO2	0.808	0.630	0.815
WBC_Count	0.3721	0.195	0.551
Diabetes (Yes)	2.8287	1.345	9.963
Cough (Yes)	7.2601	2.370	11.547
Hypertension (Yes)	6.6104	2.495	12.097
CRP	1.0272	1.004	1.042

Table 6
Credible interval finding of posterior parameters.

Variables	2.50%	25%	50%	75%	97.50%
(Intercept)	24.4444	32.23747	37.15356	41.72098	51.58415
Age (18–44)	0.5287	1.86368	2.52554	3.26089	4.96312
Age (45–64)	3.4288	4.62444	5.41182	6.20153	7.90676
Age (>65)	4.482	5.9219	6.74958	7.56156	9.35048
SpO2	−0.5124	−0.41129	−0.3613	−0.31581	−0.2351
WBC_Count	−1.7747	−1.43344	−1.23675	−1.04139	−0.72505
Diabetes (Yes)	2.5318	1.75568	1.37872	1.06351	0.40886
Cough (Yes)	3.5371	2.69687	2.24775	1.79789	1.05507
Hypertension (Yes)	3.4463	2.60355	2.20743	1.83704	1.11822
CRP	0.005	0.01708	0.02368	0.03057	0.04452

cases, which means the observed and fitted values are similar. To evaluate the model's fit, goodness of fit statistics was computed as indicated in Table 8. The Hosmer-Lemeshow statistic was not significant, indicating that there was no evidence of model is fit, and the logistic analog of R^2 stated that about 85% of the uncertainty in the presence of no surviving from Covid-19 could be of the Age, SpO2, WBC, Diabetes, Cough, Hypertension and CRP variables.

3.3. Result of Bayesian neural network

The confusion matrix for (BNN) and classic neural networks were calculated and presented below to help us identify their strengths. Following the architecture shown in Fig. 6 and Table 9, it was obvious that the Bayesian neural network approach achieved way better results with an AUC value 84.66% (95% CI of 83.26% − 85.21%) than the point estimation neural network with an AUC value 81.38% (95% CI of 80.25% − 82.03%).

3.3.1. Relative importance analysis of Bayesian neural network

To determine the contribution of each input variable in predicting death cases, the role of each input to the output is computed by multiplying the input-hidden weight with the hidden-output weight. The magnitude and direction of the connection weights determine the relative contribution, with input variables having higher connection weights representing higher intensities of signal transfer and therefore being more important than those with lower weights. To simplify the interpretation of relative importance, the attributable value of each input variable on the output is divided by the sum of contributions and expressed as a percentage, as demonstrated in Fig. 7. In comparison to the other factors, CRP, SpO2, Age, and Diabetes are the strongest predictors of increasing chances to die due to Covid-19.

3.4. Confusion Matrix analysis of applied models

It has come to the stage where all applied methods can be summarized into reasonable findings to understand and easy to follow on the model's performances. Table 10 illustrates the potential key characteristics extracted from the model after being tested with unseen datasets which was the core objective of this study. Initially, the AUC measurement is a performance evaluation metric for machine learning classification models that is calculated by dividing the number of true positives and true negatives by the total number of positive and negative observations. Bayesian neural network had the highest value with 84.66%, followed by Bayesian logistic regression with 83.07%, classic neural network with 81.38% and Logistic regression (MLE) with 80.95%. However, accuracy cannot judge the model's performance alone and there are other measurements such as F1-score, precision as well as recall.

Both neural network approaches scored the highest precision rates with 88.29% and 86.24% for the Bayesian neural network and classic neural network respectively, then followed by logistic regression (MCMC) and MLE with a success rate in predicting positive records by 85.59% and 84.55% correspondingly. Moreover, recall perspective measurement, Bayesian neural network, classic neural network, logistic regression (MCMC) and logistic regression (MLE) all had similar recall rates of close to 85%. This denotes the model's capability to appropriately foresee positives from definite positives. Unlike precision, which calculates the proportion of positive predictions made by the model that were correct, recall calculates the proportion of all positive cases that were correctly identified by the model. F1-Score was also calculated for all eight methods. It was more useful than accuracy since an uneven class distribution was present in our case. Bayesian neural network came out to be on the top of the list resulting in the highest F1-score of 87.11%. Logistic regression (MCMC) turned out to have the second-highest F1-score rate by 85.59%. This can be interpreted as the model's capacity to both catch positive cases and be precise with the cases (Table 10).

Fig. 8 displays the ROC plot of the methods where the performances can easily be detected and followed. Because ROC curves can be misleading in imbalanced datasets as in this case, precision and recall figures are frequently used instead, where the number of true positive labels is very different from the number of true negative labels.

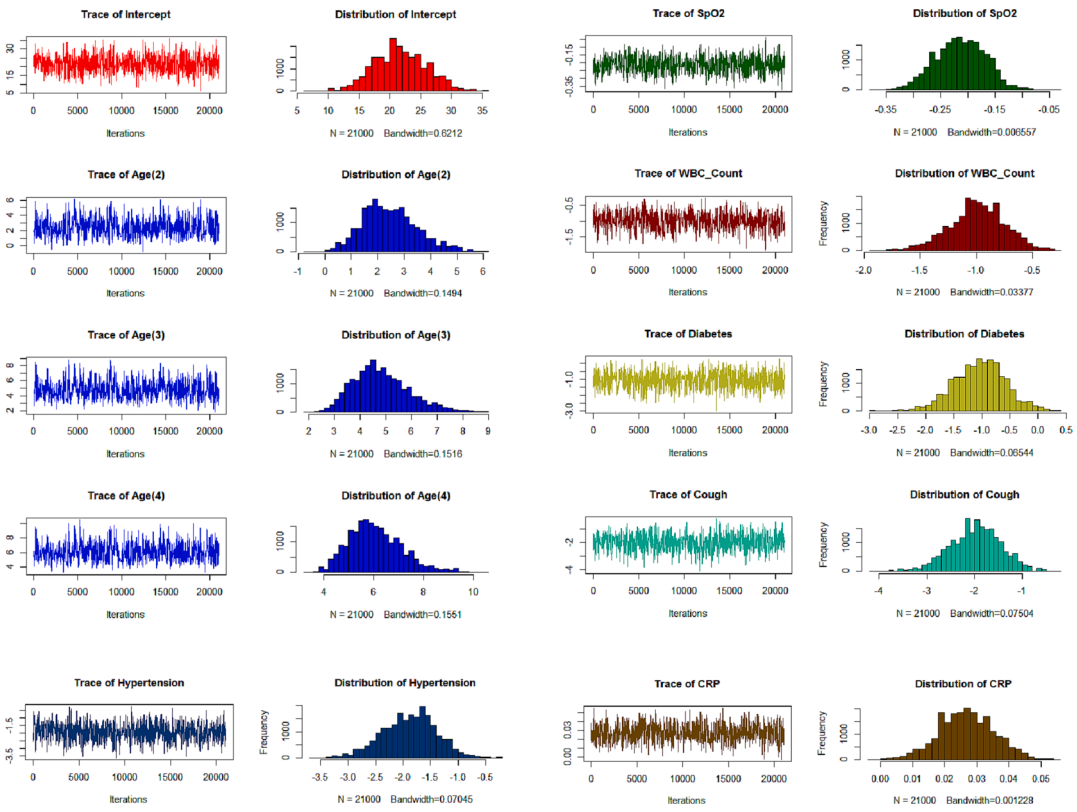

Fig. 5. Traceplots and Histogram charts of posterior distribution estimated for Bayesian logistic regression model (MCMC).

Table 7
Descriptive statistics (Minimum, Maximum, Mean, Median, SD) of the posterior findings.

Variables	Mean	Median	SD	Minimum	Maximum
(Intercept)	27.6297	21.4559	4.2894	6.1509	35.9283
Age (18–44)	2.4636	2.3519	1.0315	−0.7125	6.1421
Age (45–64)	4.8139	4.6886	1.0792	1.8271	8.8339
Age (>65)	6.0791	5.9705	1.0708	3.2534	10.4790
SpO2	−0.2132	−0.2129	0.0453	−0.3733	−0.0390
WBC_Count	−0.9886	−0.9874	0.2441	−1.9287	−0.2663
Diabetes (Yes)	1.0398	1.0305	0.4518	0.3077	2.9687
Cough (Yes)	1.9824	1.9632	0.5305	0.2308	4.1327
Hypertension (Yes)	1.8886	1.8610	0.4864	0.2638	3.6134
CRP	0.0268	0.0267	0.0085	−0.0031	0.0547

Table 8
Test of goodness-of-fit for the final model.

Statistic	Value	df	P-value
Hosmer–Lemeshow (C)	7.653	8	0.607
Deviance (G^2)	120.11	338	0.000
Nagelkerke R-Sqaure	85.21%		

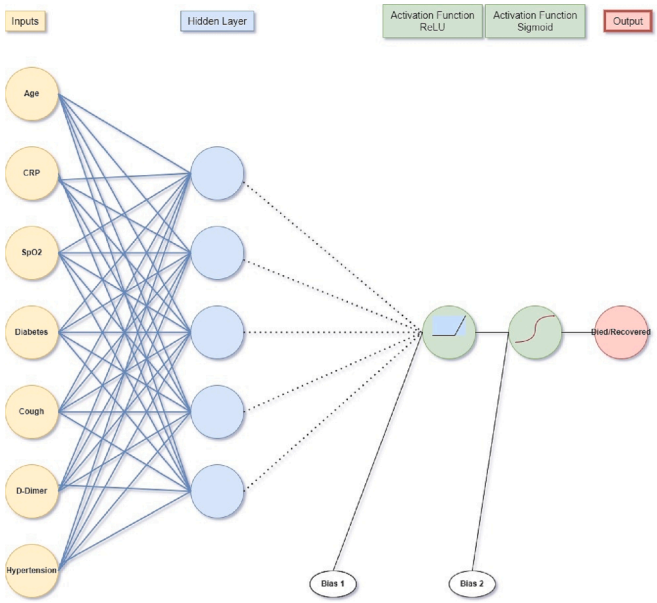

Fig. 6. Bayesian Neural Network Applied on the dataset's study.

4. Discussion

Within the context of a Bayesian methodology, the primary concentration of this work is on the integration of Bayesian neural networks, classic neural networks, Bayesian logistic regression and logistic regression (MLE). It is possible to make predictions more stable and accurate by converting the predictions into a prior distribution. Then, utilizing them as prior knowledge in the Bayesian inference process by exchanging the predictions from neural networks with predicted values for the linear regression. This will result in an improved outcome and in this instance, we applied an override Bayesian neural network and Bayesian logistic regression that can provide accurate result predictions in the presence of uncertainty. This would be the method of choice in an ideal world.

Additionally, the treatment for the Covid-19 infection is only a temporary fix utilizing the technique that is advised and after adjusting for several other variables, it was found that age had a significant correlation with the state the patients. As age increases, the odds of living become less likely to increase, and age has been recognized as the primary variable in Covid-19 patients to impact the outcome ever since the

Table 9
Confusion matrix result of Bayesian neural network and classic neural network.

	Bayesian Neural Network		Neural Network	
	Recovered	Died	Recovered	Died
Recovered	98	13	93	16
Died	16	62	19	60
AUC	0.8466		0.8138	
P-Value	0.001		0.0382	
95% CI	(0.8326, 0.8521)		(0.8025, 0.8203)	
Sensitivity	85.96%		83.04%	
Specificity	82.67%		78.95%	

beginning of the pandemic.

The early Chinese records suggest that the case-fatality rate (CRF) rises considerably beyond the age of 60, reaching 14.8%. A notable rise in the count of deceased patients was found in the patient data. According to the findings of this experiment, people in the age group (45–64) had 123 more odds to die as well as 436 times more chance in the age group 65 years old than patients who were<18 years old. It can be noticed that younger people had a higher chance to survive the disease. This is in line with previous research that has established age as a significant factor for cases that died because of Covid-19, particularly for individuals between the age of 45 and 64, and especially those over the age of 65 [40,41].

Other reports have also noted that patients in ICUs tend to be older than those who are not, and in that case fatality rates are higher among older individuals [3,42–44]. As a result, the risk of death is significantly increased in patients who are older than compared to those who are younger.

The study discovered that fever, cough, and sputum were predominant symptoms among Covid-19 individuals, mostly those who were critically ill and eventually died. It is noteworthy that cough and fever are also commonly detected as indicators in patients with severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), which are impacted by coronaviruses. Fever is recognized as a principal symptom of cytokine storms, a condition characterized by an excessive immune response and inflammation resulting from high cytokine concentrations. Moreover, Upon admission, vital signs of severely ill patients indicated elevated body temperature and respiratory rate, as well as reduced SpO2 levels. Throughout the outbreak, glucocorticoids have been placed to manage SpO2 levels below 90%, and the oxygenation saturation index is associated with both ARDS severity and increased mortality. Our result indicated that SpO2 is associated with a 20% (1 – 0.80 = 0.20) reduction in the relative risk of dying. Thus, a 1-unit increase in the corresponding oxygen level of patients admitted to the hospital due to the disease is associated with a lower risk of dying due to Covid-19.

However, the severity of Covid-19 patients was found to be unaffected by gender in our study. While initial reports from other countries suggested a higher proportion of men experiencing severe cases of Covid-19, more recent studies have shown that similar proportions of men and women are being admitted to ICUs [3,42–44], indicating that any gender differences may have diminished with the increase in incidence. It's possible that earlier reports included a higher number of males due to their higher occupational risk of infection in crowded places like markets and congregations [43]. Our study discovered that patients who passed away due to Covid-19 had more pronounced damage to white blood cells and immune cells, and the odds ratio was estimated to be 0.37; for every 1 increase in white blood cell count, an infected Covid-19 person had 63% chance to survive than not with a 95% CI of (0 and 0). Covid-19 may lead to reduced levels of T lymphocytes, including CD4 + T and CD8 + T cells, which can result in decreased production of interferon-gamma (IFN-γ), potentially contributing to disease severity [45].

Additionally, while a more intense inflammatory response was

Table 10
Confusion matrix outputs of computed models.

Classifier	AUC	F1	Precision	Recall
ANN	0.8138	0.8416	0.8532	0.8304
ANN with MCMC	0.8466	0.8711	0.8829	0.8596
Logistic Regression (MLE)	0.8095	0.8378	0.8455	0.8304
Logistic Regression (MCMC)	0.8307	0.8559	0.8559	0.8559

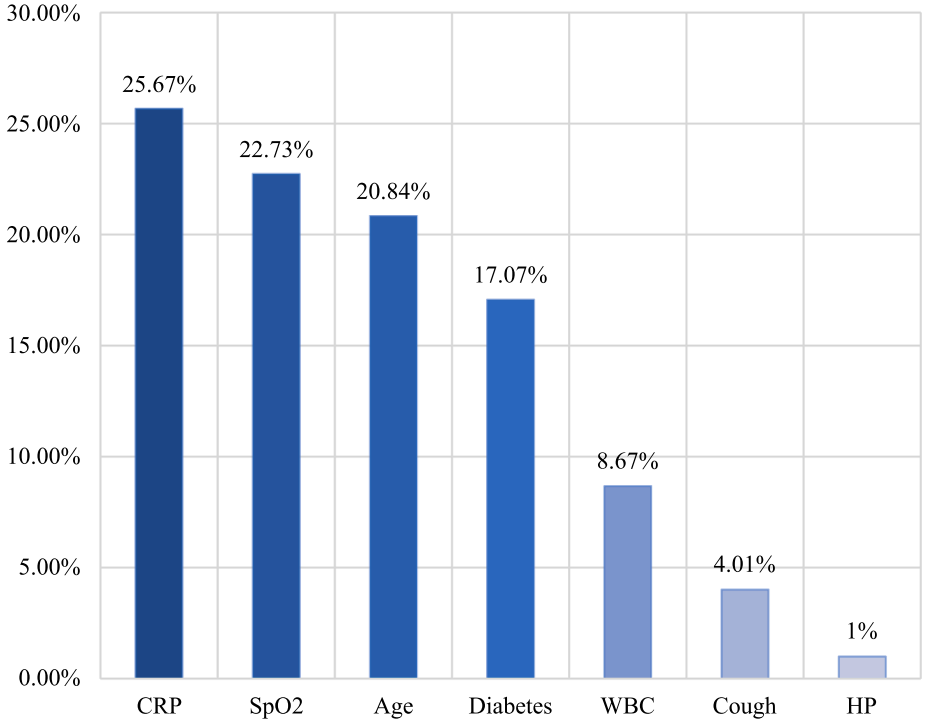

Fig. 7. Relative Importance of Inputs by Analyzing Weight Matrix.

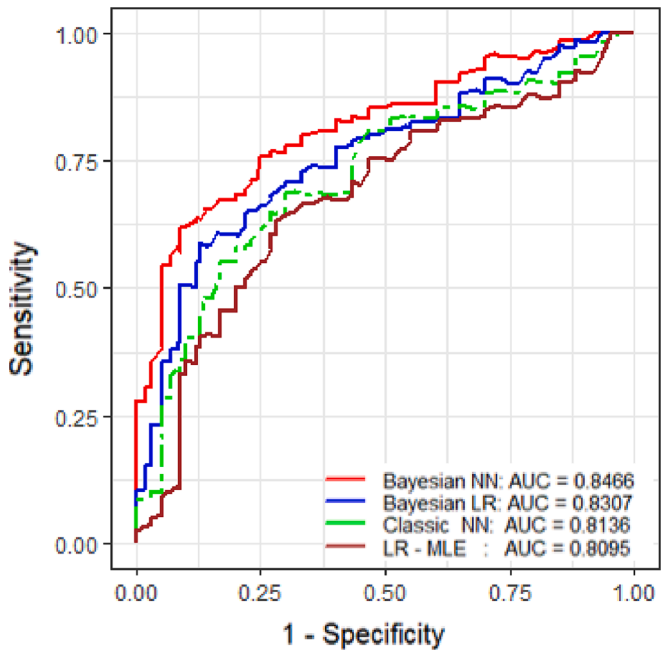

Fig. 8. ROC Evaluation Curve of all eight models applied to predict the risk of dying from Covid-19.

indicated by much higher levels of inflammatory markers, such as C-reactive protein [46,47], CRP was also one of the attributes and it was associated with a 2.72% rising chance to die. Although the effect size was small, it was statistically significant with a p-value < 0.001. D-dimer is a protein fragment that is produced in the blood after the breakdown of a blood clot through fibrinolysis [48]. In healthy individuals, D-dimer is typically not detectable in the bloodstream, except in cases where blood clots are formed and broken down. This makes a D-dimer serum test useful for ruling out thrombotic episodes and aiding in the early diagnosis of various thromboembolic conditions, including, pulmonary embolism, deep vein thrombosis, and disseminated intravascular coagulation [49,50].

Initially, it was proposed that a contact person to Covid-19 might develop a hypercoagulable state, which was supported by the formation of thromboembolisms observed in pathological studies from autopsies or biopsies [51,52]. Consequently, several researchers linked the increase in D-dimer levels to this hypercoagulable state in Covid-19 patients [53,54]. However, some researchers argue that elevated D-dimer levels may be associated with the inflammatory response rather than the thromboembolic condition in Covid-19 infected cases [55]. The exact mechanisms behind the elevated D-dimer levels in Covid-19 individuals are not yet fully understood, and further research is needed to clarify the underlying processes.

5. Limitation

Several limitations should be considered in this study. Firstly, there was a considerable amount of missing data in the laboratory and radiological records, which impeded their inclusion in the analysis. Secondly, the identified predictive factors may have been confounded by unmeasured variables, such as occupation, length of hospital stay, and pregnancy status. It is possible that medical staff and pregnant women had different disease severity profiles. Additionally, the absence of reported medications may have impacted the disease status of the infected cases and could potentially lead to different conclusions.

6. Conclusion

This research contributes by utilizing neural network approaches to learn complicated dependencies from the data as well as a Bayesian paradigm to associate the uncertainty in the predictions. Thus, our method has the potential to produce a model that can make accurate forecasts regarding infectious diseases and contribute to the mitigation of the negative effects those have diseases. To accomplish this, it is essential to keep the following information about the prior distributions in mind, as it will be utilized to estimate the parameters of the model. Even though it was assumed that these prior distributions did not provide any useful information, it is nevertheless recommended to carry out a sensitivity analysis to determine level of effectiveness.

Each parameter of the model has a normal prior distribution mean, and in addition to this, a value ranging from 102 to 106 was appended to the variance of the normal distributions. A normal prior with a variance of 106 is sufficiently non-informative and generally functions well with our dataset. This conclusion was reached as a result of the findings that the posterior distributions for the regression parameters differed only slightly from one another. This suggests that the outcomes produced by our model were reliable across a broad spectrum of prior distributions. Moreover, Bayesian neural network performed better than the other three approaches in terms of accuracy and stability as well as convergence.

7. Recommendation

The implementation of this concept is expected to make it simpler for government agencies to keep an eye out for any contagious diseases. The findings of the model can be utilized in the formulation of public health policy, such as the administration of immunizations or the implementation of preventative measures. Another suggestion is to obtain more data with much more variables such as, x-ray, MRI, medications taken by the patients.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Organization, W.H., Coronavirus disease 2019 (COVID-19): situation report, 73. 2020.
[2] X. Xie, X. Jin, G. Wei, C. Chang, Monitoring and Early Warning of SMEs' Shutdown Risk under the Impact of Global Pandemic Shock, Systems 11 (5) (2023) 260, https://doi.org/10.3390/systems11050260.
[3] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X. Zhang, L.i. Zhang, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, Lancet 395 (10223) (2020) 507–513.
[4] A.M. Dinar, E.A. Raheem, K.H. Abdulkareem, M.A. Mohammed, M.G. Oleiwie, F. H. Zayr, O. Al-Boridi, M.N. Al-Mhiqani, M.N. Al-Andoli, H.A. Khattak, Towards automated multiclass severity prediction approach for COVID-19 infections based on combinations of clinical data, Mob. Inf. Syst. 2022 (2022) 1–8.
[5] Q. Li, Y. Miao, X. Zeng, C.S. Tarimo, C. Wu, J. Wu, Prevalence and factors for anxiety during the coronavirus disease 2019 (COVID-19) epidemic among the teachers in China, Journal of Affective Disorders 277 (2020) 153–158, https://doi.org/10.1016/j.jad.2020.08.017.
[6] F. Hu, L. Qiu, X. Xi, H. Zhou, T. Hu, N. Su, Z. Duan, Has COVID-19 Changed China's Digital Trade?—Implications for Health Economics, Frontiers in public health, (2022) 10.
[7] M. Saeed, M. Ahsan, M.H. Saeed, A.U. Rahman, A. Mehmood, M.A. Mohammed, M. M. Jaber, R. Damaševičius, An optimized decision support model for COVID-19 diagnostics based on complex fuzzy hypersoft mapping, Mathematics 10 (14) (2022) 2472.
[8] K. Hameed Abdulkareem, A. Awad Mutlag, A. Musa Dinar, J. Frnda, M. Abed Mohammed, F. Hasan Zayr, A. Lakhan, S. Kadry, H. Ali Khattak, J. Nedoma, A. Yahya, Smart healthcare system for severity prediction and critical tasks management of COVID-19 patients in IoT-fog computing environments, Comput. Intell. Neurosci. 2022 (2022) 1–17.

[9] A. Onyango, B. Okelo, R. Omollo, Topological data analysis of COVID-19 using artificial intelligence and machine learning techniques in big datasets of hausdorff spaces, Journal of Data Science and Intelligent Systems (2023), https://doi.org/10.47852/bonviewJDSIS3202701.

[10] N. Luo, et al., Fuzzy logic and neural network-based risk assessment model for import and export enterprises: A review, Journal of Data Science and Intelligent Systems (2023), https://doi.org/10.47852/bonviewJDSIS32021078.

[11] F. Ahmad, Deep image retrieval using artificial neural network interpolation and indexing based on similarity measurement, CAAI Trans. Intell. Technol. 7 (2) (2022) 200–218, https://doi.org/10.1049/cit2.12083.

[12] J. Khan, E. Lee, K. Kim, A higher prediction accuracy–based alpha–beta filter algorithm using the feedforward artificial neural network, CAAI Trans. Intell. Technol. 1–16 (2022), https://doi.org/10.1049/cit2.12148.

[13] S. Sharma, K. Verma, P. Hardaha, Implementation of Artificial Intelligence in Agriculture, Journal of Computational and Cognitive Engineering (2022), https://doi.org/10.47852/bonviewJCCE2202174.

[14] Z. Chen, Research on internet security situation awareness prediction technology based on improved RBF neural network algorithm, Journal of Computational and Cognitive Engineering 1 (3) (2022) 103–108, https://doi.org/10.47852/bonviewJCCE149145205514.

[15] M. Merad, J.C. Martin, Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages, Nat. Rev. Immunol. 20 (6) (2020) 355–362.

[16] L. Carsana, A. Sonzogni, A. Nasr, R.S. Rossi, A. Pellegrinelli, P. Zerbi, R. Rech, R. Colombo, S. Antinori, M. Corbellino, M. Galli, E. Catena, A. Tosoni, A. Gianatti, M. Nebuloni, Pulmonary post-mortem findings in a series of COVID-19 cases from northern Italy: a two-centre descriptive study, Lancet Infect. Dis. 20 (10) (2020) 1135–1140.

[17] F. Wang, J. Nie, H. Wang, Q. Zhao, Y. Xiong, L. Deng, S. Song, Z. Ma, P. Mo, Y. Zhang, Characteristics of peripheral lymphocyte subset alteration in COVID-19 pneumonia, J Infect Dis 221 (11) (2020) 1762–1769.

[18] D. Wichmann, J. Sperhake, M. Lütgehetmann, et al., Autopsy findings and venous thromboembolism in patients with COVID-19: a prospective cohort study Ann Intern Med 2020, Doi 10 (2020) M20–2003.

[19] Y. Zhang, M. Xiao, S. Zhang, P. Xia, W. Cao, W. Jiang, H. Chen, X. Ding, H. Zhao, H. Zhang, C. Wang, J. Zhao, X. Sun, R. Tian, W. Wu, D. Wu, J. Ma, Y.u. Chen, D. Zhang, J. Xie, X. Yan, X. Zhou, Z. Liu, J. Wang, B. Du, Y. Qin, P. Gao, X. Qin, Y. Xu, W. Zhang, T. Li, F. Zhang, Y. Zhao, Y. Li, S. Zhang, Coagulopathy and antiphospholipid antibodies in patients with Covid-19, N. Engl. J. Med. 382 (17) (2020) e38.

[20] A. Akhmerov, E. Marbán, COVID-19 and the heart, Circ Res 126 (10) (2020) 1443–1455.

[21] M. Madjid, P. Safavi-Naeini, S.D. Solomon, O. Vardeny, Potential effects of coronaviruses on the cardiovascular system: a review, JAMA Cardiol. 5 (7) (2020) 831.

[22] L.M. Barton, E.J. Duval, E. Stroberg, S. Ghosh, S. Mukhopadhyay, Covid-19 autopsies, oklahoma, usa, Am. J. Clin. Pathol. 153 (6) (2020) 725–733.

[23] F. Colavita, D. Lapa, F. Carletti, E. Lalle, L. Bordi, P. Marsella, E. Nicastri, N. Bevilacqua, M.L. Giancola, A. Corpolongo, G. Ippolito, M.R. Capobianchi, C. Castilletti, SARS-CoV-2 isolation from ocular secretions of a patient with COVID-19 in Italy with prolonged viral RNA detection, Ann. Intern. Med. 173 (3) (2020) 242–243.

[24] I.H. Solomon, E. Normandin, S. Bhattacharyya, S.S. Mukerji, K. Keller, A.S. Ali, G. Adams, J.L. Hornick, R.F. Padera, P. Sabeti, Neuropathological features of Covid-19, N. Engl. J. Med. 383 (10) (2020) 989–992.

[25] Bernardo, J.M. and A.F. Smith, *Bayesian theory*. Vol. 405. 2009: John Wiley & Sons.

[26] E. Greenberg (Ed.), Introduction to Bayesian Econometrics, Cambridge University Press, 2012.

[27] D.H. Kadir, Likelihood approach for bayesian logistic weighted model, Cihan University-Erbil Scientific Journal 4 (2) (2020) 9–12.

[28] I. Ntzoufras (Ed.), Bayesian Modeling Using WinBUGS, Wiley, 2009.

[29] S.J. Press, S.J. Press, Bayesian statistics: principles, models, and applications, Vol. 210, John Wiley & Sons Incorporated, 1989.

[30] D.H. Kadir, K. Triantafyllopoulos, Bayesian Inference of Autoregressive Models, University of Sheffield, 2018.

[31] A. Gelman, W.R. Gilks, G.O. Roberts, Weak convergence and optimal scaling of random walk Metropolis algorithms, Ann. Appl. Probab. 7 (1) (1997) 110–120.

[32] Neal, R.M., *Bayesian learning for neural networks*. Vol. 118. 2012: Springer Science & Business Media.

[33] Betancourt, M., *A conceptual introduction to Hamiltonian Monte Carlo*. arXiv preprint arXiv:1701.02434, 2017.

[34] M.D. Hoffman, et al., Stochastic variational inference, J. Mach. Learn. Res. (2013).

[35] X. Xie, Y. Tian, G. Wei, Deduction of sudden rainstorm scenarios: integrating decision makers' emotions, dynamic Bayesian network and DS evidence theory, Natural Hazards (2022), https://doi.org/10.1007/s11069-022-05792-z.

[36] X. Xie, L. Huang, S.M. Marson, G. Wei, Emergency response process for sudden rainstorm and flooding: scenario deduction and Bayesian network analysis using evidence theory and knowledge meta-theory, Natural Hazards 117 (3) (2023) 3307–3329, https://doi.org/10.1007/s11069-023-05988-x.

[37] Hernández-Lobato, J.M. and R. Adams. *Probabilistic backpropagation for scalable learning of bayesian neural networks*. in *International conference on machine learning*. 2015. PMLR.

[38] Blundell, C., et al. *Weight uncertainty in neural network*. in *International conference on machine learning*. 2015. PMLR.

[39] D.P. Kingma, M. Welling, *An introduction to variational autoencoders*. Foundations and Trends®, Mach. Learn. 12 (4) (2019) 307–392.

[40] J. Wu, W. Li, X. Shi, Z. Chen, B. Jiang, J. Liu, D. Wang, C. Liu, Y. Meng, L. Cui, J. Yu, H. Cao, L. Li, Early antiviral treatment contributes to alleviate the severity and improve the prognosis of patients with novel coronavirus disease (COVID-19), J. Intern. Med. 288 (1) (2020) 128–138.

[41] L.E. Gralinski, V.D. Menachery, Return of the Coronavirus: 2019-nCoV, Viruses 12 (2) (2020) 135.

[42] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y.i. Hu, L.i. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L.i. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q.i. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, *China*. The lancet 395 (10223) (2020) 497–506.

[43] D. Wang, B.o. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang, Z. Peng, Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, *China*. jama 323 (11) (2020) 1061.

[44] X. Yang, Y. Yu, J. Xu, H. Shu, J. Xia, H. Liu, Y. Wu, L.u. Zhang, Z. Yu, M. Fang, T. Yu, Y. Wang, S. Pan, X. Zou, S. Yuan, Y. Shang, Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study, Lancet Respir. Med. 8 (5) (2020) 475–481.

[45] G. Chen, D.i. Wu, W. Guo, Y. Cao, D.a. Huang, H. Wang, T. Wang, X. Zhang, H. Chen, H. Yu, X. Zhang, M. Zhang, S. Wu, J. Song, T. Chen, M. Han, S. Li, X. Luo, J. Zhao, Q. Ning, Clinical and immunological features of severe and moderate coronavirus disease 2019, J. Clin. Invest. 130 (5) (2020) 2620–2629.

[46] P. Jousilahti, et al., The association of c-reactive protein, serum amyloid a and fibrinogen with prevalent coronary heart disease—baseline findings of the PAIS project, Atherosclerosis 156 (2) (2001) 451–456.

[47] N.R. Sproston, J.J. Ashworth, Role of C-reactive protein at sites of inflammation and infection, Front. Immunol. 9 (2018) 754.

[48] B. Jacobs, A. Obi, T. Wakefield, Diagnostic biomarkers in venous thromboembolic disease, J. Vasc. Surg. Venous Lymphat. Disord. 4 (4) (2016) 508–517.

[49] G. Le Gal, H. Bounameaux, D-dimer for the diagnosis of pulmonary embolism: a call for sticking to evidence, Intensive Care Med 31 (1) (2005) 1–2.

[50] J.K. Schaefer, B. Jacobs, T.W. Wakefield, S.L. Sood, New biomarkers and imaging approaches for the diagnosis of deep venous thrombosis, Curr. Opin. Hematol. 24 (3) (2017) 274–281.

[51] Z. Lang, et al., Pathological study on severe acute respiratory syndrome, Chin Med J (Engl) 116 (07) (2003) 976–980.

[52] Z. Xu, L. Shi, Y. Wang, J. Zhang, L. Huang, C. Zhang, S. Liu, P. Zhao, H. Liu, L. i. Zhu, Y. Tai, C. Bai, T. Gao, J. Song, P. Xia, J. Dong, J. Zhao, F.-S. Wang, Pathological findings of COVID-19 associated with acute respiratory distress syndrome, Lancet Respir. Med. 8 (4) (2020) 420–422.

[53] F.L. Wright, T.O. Vogler, E.E. Moore, H.B. Moore, M.V. Wohlauer, S. Urban, T. L. Nydam, P.K. Moore, R.C. McIntyre, Fibrinolysis shutdown correlation with thromboembolic events in severe COVID-19 infection, J. Am. Coll. Surg. 231 (2) (2020) 193–203.

[54] F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu, L. Guan, Y. Wei, H. Li, X. Wu, J. Xu, S. Tu, Y.i. Zhang, H. Chen, B. Cao, Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study, Lancet 395 (10229) (2020) 1054–1062.

[55] B. Yu, X. Li, J. Chen, M. Ouyang, H. Zhang, X. Zhao, L. Tang, Q. Luo, M. Xu, L. Yang, G. Huang, X. Liu, J. Tang, Evaluation of variation in D-dimer levels among COVID-19 and bacterial pneumonia: a retrospective analysis, J. Thromb. Thrombolysis 50 (3) (2020) 548–557.