PREDICTION OF NOVEL CORONAVIRUS-19 BASED ON EXTREME LEARNING MACHINE ALGORITHM

SM Saravanakumar[1]*, Dr. T. Revathi[2]
[1]Assistant Professor & Research Scholar, Department of Computer Science, PSG College of Arts & Science, Coimbatore , India
[2]Department of Computer Science, PSG College of Arts & Science, Coimbatore , India
***Corresponding Author**

Abstract

In March 2020, the World Health Organization (WHO) declared the new corona virus pneumonia (COVID-19) as a world pandemic, which means that the epidemic has broken out worldwide. COVID-19 is a highly contagious virus which almost freezes the world along with its economy. Its ability of human-to-human and surface-to-human transmission turns the world into catastrophic phase. The main clinical symptoms of COVID-19 are fever, cough, and fatigue, which may lead to a fatal complication: acute respiratory distress syndrome. The main challenge in inhibiting the spread of this disease is the lack of efficient detection methods. In clinical practice, by combining clinical symptoms and travel history, CT Scan is an efficient and safe method to diagnose COVID-19. Computer Aided Diagnostic technology improves the sensitivity and specificity of doctor's diagnosis and is accurate and efficient, which helps rapid diagnosis of a large number of suspicious cases. For example, the out preformed AI-assisted diagnosis system can achieve an accuracy rate comparable to that of radiologists, and it take less than 1 second to perform a diagnosis. In this research work, the main aim is to predict the future conditions of COVID-19 to recede its impact. The prediction results show the superiority of the proposed intelligent predictors with accuracy greater than 98%. Therefore, medical personnel can take defensive steps earlier.

Keywords: **Machine Learning Algorithms, Coronavirus Disease 2019 (COVID-19), Disease Prediction, SARS-CoV-19.**

DOI Number:10.14704/nq.2022.20.8.NQ44599 **NeuroQuantology 2022; 20(8): 5714-5726**

1. INTRODUCTION

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has spread worldwide and has infected 119.21 million people and killed 2.64 million people as of 14 March 2021 [1]. The increasing number of infections is putting enormous pressure on health services worldwide, and how to use the simplest indicators to identify and give early intervention to COVID-19 patients who are at risk of death or require mechanical ventilation for endotracheal intubation has become an urgent problem.Several studies have been conducted to predict in-hospital death in patients with COVID-19 [2-8], and the inclusion metrics and methodologies of these studies vary (Table 1), such as Kaufmann et al. [2]. A predictive model was constructed using adal natriuretic peptides at the time of admission of COVID-19 patients to predict patient mortality, and the AUC value of the ROC curve was 0.832; Zhang et al. [3] Clinical indicators and chest CT performance of COVID-19 patients were used to assess patient prognosis. However, the existing predictive models mainly have the following problems: (1) the sample size used to build the predictive model in some studies is not large; (2) the predictors determined by some

studies cannot be quickly obtained in the clinic.

Table 1. Literature review of predicting in-hospital mortality in patients with COVID-19

Study	Predictive factor	Method	*n*	AUC
Kaufmann, et al[2]	Mid-regional pro-ANP	Cox regression	213	0.832
Zhang, et al[3]	CT features, age, LDH, diarrhea	XGBoost	198	0.924
Zhang, et al [4]	D-dimer	Cox proportional hazard model	343	0.890
Bertsimas, et al[5]	Age, oxygen saturation, CRP, BUN, creatinine	XGBoost	3 927	0.920
Zhao, et al[6]	Heart failure, procalcitonin, LDH, COPD, oxygen saturation, heart rate, age	Logistic regression model	641	0.830
Peng, et al[7]	NLR, LDH, IL-6, age, PaO_2/FiO_2	Logistic regression model	1 477	0.910
Li, et al[8]	Age, severity at admission, dyspnea, cardiovascular disease, LDH, TBil, glucose, urea	Logistic regression model	4 086	0.920

COVID-19: Coronavirus disease 2019; AUC: Area under curve; pro-ANP: Pro-atrial natriuretic peptide; CT: Computed tomography; LDH: Lactate Dehydrogenase; CRP: C Reactive Protein; BUN: Blood Urea Nitrogen; COPD: Chronic Obstructive Pulmonary Disease; NLR: Neutrophil-to-Lymphocyte Ratio; IL-6: Interleukin 6; PaO_2: Arterial partial pressure of oxygen; FiO_2: Fraction of inspired oxygen; TBil: Total Bilirubin.

This research work mainly used the routine examination indicators of large samples of covid-19 patients to construct a prediction model using Extreme Learning Machine (ELM) algorithm, to predict the clinical outcome of patients, so as to help clinicians quickly assess the prognosis of covid-19 patients with routine examination indicators and intervene in a timely manner.

2. MATERIALS AND METHODS

2.1 Dataset

In this paper, datasets of India confirmed and death cases are being used. Datasets of India are taken from the Ministry of Health and Family Welfare, Government of India [9]. Datasets used in this research to carry out the experimental predictive analysis with the categorization of data is shown in Table 3. Randomly 4804 PATIENTS WITH COVID-19 WHO WERE TREATED for Covid-19 were taken. All selected patients were diagnosed with COVID-19 according to the Diagnosis and Treatment Plan for Pneumonia Infected by Novel Coronavirus [9] issued by the ICMR (Indian Council for Medical Research), admission-related laboratory test results, comorbidities and clinical outcomes were extracted from the patient's electronic medical records, and the proportion of patients with missing information was excluded > 20%.

2.2 Feature Extraction and Selection

The work included patient demographic information, such as age and sex; previous disease history, such as hypertension, diabetes, coronary heart disease, chronic obstructive pulmonary disease, kidney disease, malignancy; basic examination indicators at first admission, such as blood pressure, respiratory rate, pulse rate; clinical symptoms, such as cough, fatigue, fever, sputum cough, wheezing; blood routine examination at first admission, such as white blood cell count, absolute lymphocyte value, monocyte absolute value, neutrophil absolute value, eosinophil absolute value, basophil absolute value, red blood cell count, haemoglobin, haematocrit, mean red blood cell volume, mean red blood cell haemoglobin

amount, platelet count, average platelet volume, hypersensitivity CRP; admission liver function tests, such as alanine aminotransferase, aspartate aminotransferase, alkaline phosphatase, total protein, albumin, urea, total bilirubin, direct bilirubin, carbon dioxide binding, total bile acid; admission to the hospital kidney function test, such as serum creatinine, blood urea nitrogen, uric acid, sodium ion, potassium ion, calcium ion, chloride ion, etc.

2.3 Clinical Outcomes

Adverse clinical outcome measures, including in-hospital death and nosocomial intubation, were obtained from clinical records.

2.4 Statistical Processing

Extreme Learning Machine (ELM) algorithm were used to model the prediction of clinical outcomes in patients with COVID-19 and compared with Support Vector Machine (SVM),Artificial Neural Network (BPN), Naive Bayes (NB), Logistic Regression (LR), and Random Forest (RF). The partial patient data is used as the training set, and the remaining patient data is used as the test set. Due to an imbalance between positive and negative values of the outcome variables in the training set (2972 survivors, 68 in-hospital deaths; 2982 without endotracheal intubation and 58 with endotracheal intubation), oversampling was used to balance the ratio of positive to negative [10-11].Finally, the proportion of nosocomial deaths was converted from 43.7:1 to 50:50, and the sample size after oversampling was 5331, and the proportion of whether they received endotracheal intubation was changed from 51.4:1 to 50:50, and the sample size after oversampling was 5922, and the main process was shown in Figure 1.

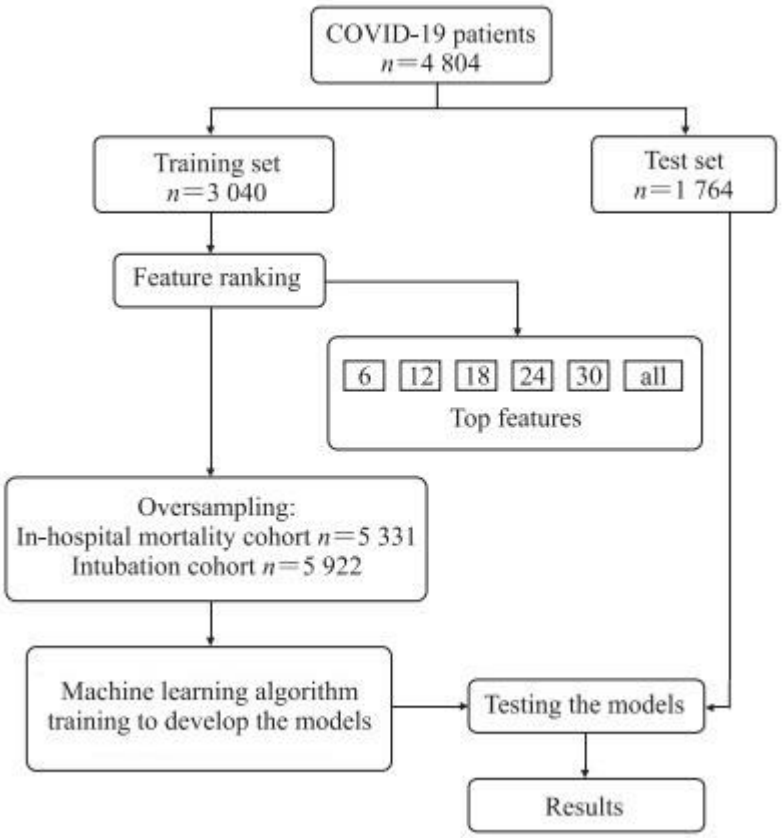

Figure 1. Flow chart of data training and testingCOVID-19;

Imbalanced classification is a supervised form of learning [10]. The proportion of one classification in the data it processes is much larger than that of other classifications, which is especially common in binary classification data. when working with unbalanced classification data, the inability to obtain enough information from a small sample size

classification can cause instability in the algorithm and lead to biased predictions. the oversampling method is a method of processing unbalanced data, using the oversampling method to generate too many samples of the classification according to the law of less classified samples, so that the data tend to be balanced.

The Extreme Learning Machine (ELM) proposed by Huang et al. [11] is a model based on a single hidden layer neural network, which after setting the weights and biases of the hidden layer, and based on a single layer feed forward network (SLFN) algorithm, allows to obtain a closed form solution for the output weights by a least squares solution. It is drawn from a continuous probability distribution function. The advantages of the ELM model are, (1) Less complex design; (2) High classification accuracy; (3) Good generalization ability; (4) Less computing time. During training process, the weights and biases of the hidden Single hidden Layer Feed-forward Networks (SLFN) are randomly adopted and never updated. The output weights are stated using the generalized Moore-Penrose generalized inverse of the hidden layer output matrix.

The ANN algorithm is similar to the biological nervous system and relies on nodes and connections to work [12]. The ANN algorithm reduces the error value of the loss function by adjusting the weights by the backpropagation algorithm, which is the standard process of supervised learning algorithms. The Naive Bayes algorithm is a simple classifier based on Baye's theorem.[13] The algorithm first estimates the conditional probability of each variable from the training set, and then continuously corrects the probability according to the actual results, and finally finds the output with the largest posterior probability. Logistic regression algorithm is a regression model based on the probability of occurrence of an event as the dependent variable and the influencing factors of the dependent variable as the independent variable, and the algorithm predicts the probability of various types through the sigmoid function. The random forest algorithm is a classifier that uses multiple trees to train and predict samples, consisting of several independent decision trees.[14]

This research work used a feature ranking method to identify the variables that had the most significant impact on clinical outcomes. The eigenvalues are sorted in 6/12/18/24/30/all groups, and the optimal combination of variables is selected using recursive feature elimination. When training the dataset, the parameter settings of the four machine learning algorithms are all 100% repeated cross-validation, which is repeated 3 times. All of these machine learning algorithms are evaluated by sensitivity, specificity, accuracy, and AUC values.

This research work was statistically analysed using MATLAB tool under windows environment. The normally distributed measurement data are expressed as $x \pm s$, and the skewed distribution is expressed as the median (lower quartile, upper quartile), and the independent sample *t-test* and the Mann-Whitney *U* test are used to compare the intergroup data, respectively. The counting data are expressed as the number of examples and percentages, and the intergroup data are compared using the $\chi 2$ test or fisher exact probability method. Missing values in the data are handled using multiple imputation (see Table 2 for missing data). The test level (α) is 0.05.

Table 2. Missing data of whole group of cases

N = 4804, *n* (%)				
Variable	Missing		Variable	Missing
WBC	54 (1.12)		AST	140 (2.91)

Lymphocyte	54 (1.12)	ALP	146 (3.04)
Monocyte	54 (1.12)	TP	144 (3.00)
Neutrophil	54 (1.12)	Albumin	145 (3.02)
Eosinophil	54 (1.12)	TBil	145 (3.02)
Basophil	55 (1.14)	DBil	158 (3.29)
RBC	54 (1.12)	CO_2	278 (5.79)
Hemoglobin	54 (1.12)	TBA	177 (3.68)
Hematocrit	54 (1.12)	Na^+	281 (5.85)
MCV	54 (1.12)	K^+	286 (5.95)
MCH	54 (1.12)	Ca^{2+}	286 (5.95)
Platelet	55 (1.14)	Cl^-	277 (5.77)
MPV	54 (1.12)	Creatinine	205 (4.27)
hs-CRP	397 (8.26)	UA	209 (4.35)
ALT	150 (3.12)	CK-MB	824 (17.15)

WBC: White blood cell; RBC: Red blood cell; MCV: Mean corpuscular volume; MCH: Mean corpuscular hemoglobin; MPV: Mean platelet volume; hs-CRP: Hypersensitivity C reactive protein; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; ALP: Alkaline phosphatase; TP: Total protein; TBil: Total bilirubin; DBil: Direct bilirubin; TBA: Total bile acid; UA: Uric acid; CK-MB: Creatine kinase-myocardial band.

3. EXPERIMENTAL RESULTS

3.1 BASELINE FEATURES IN PATIENTS WITH COVID-19

Total of the 4804 patients with COVID-19, 87 (1.81%) received endotracheal intubation during hospitalization, 4717 did not undergo endotracheal intubation during hospitalization, 100 (2.08%) died before discharge and 4704 survived discharge. Analysis of baseline characteristics of patients with COVID-19 (Table 3) showed that nosocomial deaths were older than patients undergoing endotracheal intubation, with a higher proportion of men, faster respiratory rate upon admission, lower diastolic blood pressure, more diabetes mellitus, coronary heart disease, chronic obstructive pulmonary disease, history of kidney disease, white blood cell count, absolute neutrophils, average platelet volume, hypersensitivity CRP, Aspartate aminotransferase, alkaline phosphatase, urea, total bilirubin, direct bilirubin, serum creatinine, blood urea nitrogen, cardio myocardial creatine kinase isoenzyme levels were higher, while the absolute values of lymphocytes, absolute values of eosinophils, absolute values of basophils, red blood cell count, haemoglobin, haematocrit, platelet count, total protein, albumin, uric acid, calcium ion levels were low, and the differences were statistically significant (*P* The average <0.05); in addition, compared with patients with no adverse clinical outcomes, the patients who died in the hospital had fewer cough symptoms, lower absolute values of monocytes and

lower levels of <b12>CO2, and patients who received endotracheal intubation in the hospital had higher levels of alanine aminotransferase, lower serum chloride ion levels, and more people with a history of malignant tumours, and the difference was statistically significant (P <0.05).

Table 3. Baseline characteristics of COVID-19 patients with different outcomes

Characteristic	In-hospital mortality		In-hospital tracheal intubation	
	Alive N = 4 704	Died N = 100	Non-intubation N = 4 717	Intubation N = 87
General condition				
Age, *n* (%)				
＜18 years	21 (0.45)	0	21 (0.45)	0
18-44 years	900 (19.13)	1 (1.00)	901 (19.10)	0
45-59 years	1 507 (32.04)	12 (12.00)	1507 (31.95)	12 (13.79)
≥60 years	2 276 (48.38)	87 (87.00)	2 288 (48.51)	75 (86.21)
Male, *n* (%)	2 220 (47.19)	66 (66.00)	2 226 (47.19)	60 (68.97)
RR/min^{-1}, M (Q_L, Q_U)	20.0 (19.0, 20.0)	21.0 (20.0, 24.0)	20.0 (19.0, 20.0)	21.0 (19.0, 23.0)
Pulse/min^{-1}, M (Q_L, Q_U)	84.0 (78.0, 94.0)	88.0 (78.0, 98.0)	84.0 (78.0, 94.0)	88.0 (78.0, 98.0)
SBP/mmHg, M (Q_L, Q_U)	130 (120, 140)	130 (120, 144)	130 (120, 140)	131 (120, 143)
DBP/mmHg, M (Q_L, Q_U)	80 (75, 89)	76 (67, 86)	80 (75, 89)	78 (68, 86)
Clinical symptom, *n* (%)				
Cough	2 717 (57.76)	45 (45.00)	2 717 (57.60)	45 (51.72)
Fatigue	1 651 (35.10)	38 (38.00)	1 657 (35.13)	32 (36.78)
Fever	2 917 (62.01)	61 (61.00)	2 924 (61.99)	54 (62.07)
Sputum	243 (5.17)	4 (4.00)	243 (5.15)	4 (4.60)
Gasp	719 (15.28)	20 (20.00)	722 (15.31)	17 (19.54)
Laboratory test, M (Q_L, Q_U)				
WBC/(L^{-1}, $\times10^9$)	5.70 (4.70, 6.90)	8.50 (6.45, 12.10)	5.70 (4.70, 6.90)	8.80 (6.65, 12.80)
Lymphocyte/(L^{-1}, $\times10^9$)	1.53 (1.16, 1.90)	0.64 (0.42, 0.93)	1.53 (1.16, 1.90)	0.74 (0.42, 1.02)
Monocyte/(L^{-1}, $\times10^9$)	0.41 (0.32, 0.52)	0.36 (0.24, 0.56)	0.41 (0.32, 0.52)	0.36 (0.24, 0.60)
Neutrophil/(L^{-1}, $\times10^9$)	3.44 (2.67, 4.47)	7.25 (5.30, 10.7)	3.44 (2.67, 4.47)	7.25 (5.35, 11.90)
Eosinophil/(L^{-1}, $\times10^9$)	0.11 (0.07, 0.19)	0.02 (0.01, 0.06)	0.11 (0.07, 0.19)	0.02 (0.01, 0.06)

Characteristic	In-hospital mortality		In-hospital tracheal intubation	
	Alive $N = 4704$	Died $N = 100$	Non-intubation $N = 4717$	Intubation $N = 87$
Basophil/(L^{-1}, $\times 10^9$)	0.02 (0.01, 0.03)	0.01 (0.01, 0.01)	0.02 (0.01, 0.03)	0.01 (0.01, 0.02)
RBC/(L^{-1}, $\times 10^{12}$)	4.07 (3.73, 4.41)	3.90 (3.39, 4.23)	4.07 (3.73, 4.41)	3.96 (3.37, 4.25)
Hemoglobin/($g \cdot L^{-1}$)	126.0 (116.0, 137.0)	117.0 (102.0, 132.0)	126.0 (116.0, 137.0)	121.0 (104.0, 134.0)
Hematocrit/%	37.30 (34.50, 40.30)	35.20 (31.10, 38.50)	37.30 (34.50, 40.30)	36.00 (31.40, 39.30)
MCV/fL	92.30 (89.60, 94.90)	92.30 (88.00, 96.70)	92.30 (89.60, 94.90)	92.50 (89.00, 97.20)
MCH/pg	31.20 (30.20, 32.20)	31.00 (29.70, 32.60)	31.20 (30.20, 32.20)	31.40 (29.80, 32.50)
Platelet/(L^{-1}, $\times 10^9$)	221.0 (181.0, 270.0)	152.0 (88.2, 230.0)	221.0 (181.0, 270.0)	159.0 (96.2, 249.0)
MPV/fL	9.80 (9.20, 10.60)	10.50 (9.75, 11.40)	9.90 (9.20, 10.60)	10.30 (9.65, 11.40)
hs-CRP/($mg \cdot L^{-1}$)	1.59 (0.63, 3.77)	12.10 (6.26, 84.60)	1.59 (0.63, 3.77)	16.10 (6.66, 84.00)
ALT/($U \cdot L^{-1}$)	21.60 (14.00, 36.10)	24.00 (15.70, 44.60)	21.60 (14.00, 36.00)	26.00 (16.90, 49.80)
AST/($U \cdot L^{-1}$)	18.40 (14.40, 25.10)	30.60 (21.50, 47.40)	18.40 (14.40, 25.10)	27.10 (19.80, 42.20)
ALP/($U \cdot L^{-1}$)	69.80 (58.00, 84.00)	89.60 (68.80, 117.00)	69.80 (58.00, 83.90)	91.20 (69.00, 116.00)
TP/($g \cdot L^{-1}$)	67.00 (62.70, 71.50)	60.20 (55.40, 65.40)	67.00 (62.60, 71.40)	61.00 (56.10, 66.10)
Albumin/($g \cdot L^{-1}$)	38.30 (35.40, 40.70)	30.50 (27.10, 33.40)	38.30 (35.40, 40.70)	31.40 (27.50, 34.90)
Creatinine/($\mu mol \cdot L^{-1}$)	64.10 (55.00, 75.60)	76.40 (62.40, 105.00)	64.10 (55.10, 75.80)	73.80 (58.90, 90.90)
TBil/($\mu mol \cdot L^{-1}$)	9.40 (7.30, 12.40)	13.00 (9.55, 19.50)	9.40 (7.30, 12.40)	12.60 (10.10, 18.90)
DBil/($\mu mol \cdot L^{-1}$)	3.50 (2.60, 4.70)	6.90 (4.47, 10.50)	3.50 (2.60, 4.70)	6.45 (4.43, 9.52)
CO_2/($mmol \cdot L^{-1}$)	24.10 (22.70, 25.60)	22.10 (19.60, 25.90)	24.10 (22.70, 25.60)	23.60 (20.60, 26.60)
TBA/($\mu mol \cdot L^{-1}$)	3.80 (2.40, 6.10)	4.30 (2.58, 6.82)	3.80 (2.40, 6.10)	4.00 (2.70, 6.88)

Characteristic	In-hospital mortality		In-hospital tracheal intubation	
	Alive $N=4$ 704	Died $N=100$	Non-intubation $N=4$ 717	Intubation $N=87$
BUN/(mmol·L^{-1})	4.41 (3.62, 5.46)	8.46 (5.29, 12.40)	4.41 (3.62, 5.48)	6.66 (4.97, 11.20)
UA/(μmol·L^{-1})	284.0 (231.0, 347.0)	262.0 (192.0, 337.0)	284.0 (231.0, 348.0)	230.0 (176.0, 319.0)
Na^+/(mmol·L^{-1})	141.0 (139.0, 143.0)	140.0 (137.0, 145.0)	141.0 (139.0, 143.0)	140.0 (136.0, 144.0)
K^+/(mmol·L^{-1})	4.18 (3.90, 4.49)	4.30 (3.80, 4.60)	4.18 (3.90, 4.49)	4.30 (3.74, 4.60)
Ca^{2+}/(mmol·L^{-1})	2.17 (2.10, 2.24)	1.97 (1.87, 2.03)	2.17 (2.10, 2.23)	1.97 (1.88, 2.05)
Cl^-/(mmol·L^{-1})	106.0 (104.0, 108.0)	104.0 (99.0, 110.0)	106.0 (104.0, 108.0)	104.0 (99.4, 108.0)
CK-MB/(U·L^{-1})	7.30 (1.20, 9.70)	12.80 (8.10, 22.50)	7.30 (1.20, 9.80)	10.30 (5.62, 17.30)
Medical history, *n* (%)				
Hypertension	1 429 (30.38)	28 (28.00)	1 429 (30.29)	28 (32.18)
Diabetes	630 (13.39)	25 (25.00)	634 (13.44)	21 (24.14)
CHD	269 (5.72)	22 (22.00)	276 (5.85)	15 (17.24)
COPD	43 (0.91)	6 (6.00)	42 (0.89)	7 (8.05)
Kidney disease	124 (2.64)	16 (16.00)	126 (2.67)	14 (16.09)
Cancer	42 (0.89)	3 (3.00)	41 (0.87)	4 (4.60)

1 mmHg = 0.133 kPa. COVID-19: Coronavirus disease 2019; RR: Respiratory rate; SBP: Systolic blood pressure; DBP: Diastolic blood pressure; WBC: White blood cell; RBC: Red blood cell; MCV: Mean corpuscular volume; MCH: Mean corpuscular hemoglobin; MPV: Mean platelet volume; hs-CRP: Hypersensitivity C reactive protein; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; ALP: Alkaline phosphatase; TP: Total protein; TBil: Total bilirubin; DBil: Direct bilirubin; TBA: Total bile acid; BUN: Blood urea nitrogen; UA: Uric acid; CK-MB: Creatine kinase-myocardial band; CHD: Coronary heart disease; COPD: Chronic obstructive pulmonary disease; *M* (Q_L, Q_U): Median (lower quartile, upper quartile).

3.2 Machine Learning Algorithms Predict Clinical Outcomes

(1) Prediction of Death in Hospital: The variables most associated with in-hospital death were white blood cell count, albumin, calcium ions, blood urea nitrogen, myocardial creatine kinase isoenzyme, and age. These 6 variables were incorporated into four machine learning algorithms to build the model, and the four prediction models in the test set showed good results (Table 4),the AUC value of the ELM algorithm was 0.994 (98% *CI* 0.874~0.100), the AUC value of the SVM algorithm was 0.972 (96% *CI* 0.862~0.990), the AUC value of the ANN algorithm was 0.938 (95% *CI* 0.882~0.993), and the AUC value of the Naive Bayes algorithm was 0.952 (95% *CI* 0.925~0.979), logistic regression algorithm has an AUC value of 0.926

(95% *CI* 0.865~0.987), and the AUC value of the stochastic forest algorithm is 0.867 (95% *CI* 0.780~0.954), of which the prediction model built with the naïve Bayes algorithm has the best performance.

Table 4. Predictive Performance of Prediction Models for the Test Data

Outcome	Sensitivity	Specificity	Accuracy
In-hospital Mortality			
ELM	0.832	0.957	**0.968**
SVM	0.824	0.953	0.967
BPN	0.812	0.954	0.952
NB	0.750	0.967	0.963
LR	0.750	0.954	0.951
RF	0.332	0.946	0.945
In-hospital Tracheal Intubation			
ELM	0.904	0.899	**0.941**
SVM	0.871	0.898	0.896
BPN	0.844	0.897	0.896
NB	0.938	0.887	0.888
LR	0.844	0.915	0.914
RF	0.406	0.976	0.902

COVID-19: Coronavirus disease 2019; ELM: Extreme Learning Machine; SVM: Support Vector Machine; ANN: Artificial neural network; NB: Naive Bayes; LR: Logistic regression; RF: Random forest.

(2) Prediction of Endotracheal Intubation in the Hospital: The variables most associated with in-hospital endotracheal intubation were white blood cell count, absolute lymphocyte value, hypersensitivity CRP, total bilirubin, calcium ions, and age. These six variables were incorporated into four machine learning algorithms to build the model, and the four prediction models in the test set showed good results (Table 4), the AUC value of the ELM algorithm was 0.982 (98% *CI* 0.800~0.100), the AUC value of the SVM algorithm was 0.968 (96% *CI* 0.844~0.990), the AUC value of the ANN algorithm was 0.932 (95% *CI* 0.814~0.980), and the AUC value of the Naive Bayes algorithm was 0.948 (95% *CI* 0.896~0.965), logistic regression algorithm has an AUC value of 0.935 (95% *CI* 0.817~0.981), and the AUC value of the random forest algorithm is 0.936 (95% *CI* 0.921~0.972), of which the prediction model built with ELMalgorithm has the best performance.

4. FINDINGS AND DISCUSSION

In this research, extreme learning machine algorithm were used to construct a prediction model for the clinical outcomes of covid-19 patients. The test results showed that the constructed model performed well in predicting patients' nosocomial death and whether they received endotracheal intubation. The results of this research work also showed that the combination of six features, white blood cell count, albumin, calcium ions, blood urea nitrogen, myocardial creatine kinase isoenzyme, and age, was the best combination of variables to predict nosocomial death in patients with covid-19, and that the combination of six features, including white blood cell count, absolute lymphocyte value, hypersensitivity CRP, total bilirubin, calcium ions, and age, was the best combination of variables to predict whether

patients with covid-19 would receive endotracheal intubation.

Li et al. [8] Predicted nosocomial death in patients with COVID-19 in the same batch of this research work using logistic regression, including age, severity of illness upon admission, dyspnea, cardiovascular disease, lactate dehydrogenase, total bilirubin, blood glucose, and urea, and the predictor model constructed performed well in the external test set (AUC value of 0.920, 95% CI). 0.86~0.98. This research work took in-hospital death in patients with COVID-19 as the primary clinical outcome and endotracheal intubation as the secondary clinical outcome, while focusing on the first laboratory examination indicators of admission to the hospital of patients with COVID-19, and used a variety of machine learning algorithms to compare.

Serum albumin is the most important protein in plasma and maintains stable nutrient and plasma colloidal osmolality in the human body. Albumin levels are closely related to inflammatory status, and the body's inflammatory response can increase capillary permeability, leading to serum albumin escape [15]. In addition, as the interstitial space expands, the volume of albumin distribution increases, resulting in a shortened albumin half-life. Exacerbation of the condition in patients with COVID-19 may trigger severe inflammation. Li et al. [16] In a study of 523 patients with COVID-19, hospital stay was significantly shortened in patients with reduced albumin levels. Previous studies have reported an independent association of albumin with mortality in patients with COVID-19 [17]. This research work also found a strong correlation between serum albumin levels and nosocomial death in patients with COVID-19.

Studies have shown that calcium ions can cause a cellular inflammatory response by impairing mitochondrial function [18]. The cellular inflammatory response is thought to be associated with death in patients with COVID-19 [19]. Cell experimental results have shown that calcium channel blockers are able to prevent replication of SARS-CoV-2 [20]. A multicenter study by 39 hospitals showed that calcium channel blockers were effective in reducing mortality in patients with COVID-19 [21]. The results of this research work show a strong correlation between calcium ions and nosocomial death and endotracheal intubation in patients with COVID-19.

Age is considered an independent risk factor for severe illness and death in patients with COVID-19 [19, 22]. Elderly patients are often accompanied by cardiovascular and cerebrovascular diseases, which may worsen the symptoms of COVID-19. The results of this research work show that age is associated with both nosocomial death and endotracheal intubation in patients with COVID-19.

Acute heart injury is the most reported cardiovascular disease in patients with COVID-19, with an incidence of 8 to 12 percent; direct myocardial injury and systemic inflammation caused by viral invasion of cardiomyocytes appear to be the most common mechanisms for heart damage [23]. When myocardial tissue damage is severe, myocardial creatine kinase isoenzyme is released into the bloodstream, and elevated myocardial creatine kinase isoenzyme in the serum becomes an important criterion for diagnosing acute heart injury. A study based on 273 patients with COVID-19 conducted by Han et al. [24] found that elevated concentrations of myocardial creatine kinase isoenzymes were associated with the severity of COVID-19. This research work also confirms this conclusion.

White blood cells are a very important class of blood cells in the human body, with the ability to engulf foreign bodies and produce antibodies. Elevated white blood cell count is more common in inflammation, infection, etc. The white blood cell count in patients with severe and dead COVID-19 is significantly elevated compared to non-severe patients [25]. Studies have shown that elevated white blood cell counts are an independent risk factor for death in patients with COVID-19 [26], and the results of this research work are consistent with this.

Blood urea nitrogen is the main end product of protein metabolism. Previous studies have found that blood urea nitrogen can be used as an indicator of organ failure after 48 hours of hospital admission [27], and it also has a role in assessing renal function. Elevated blood urea nitrogen levels are also a predictor of a poorer prognosis in patients with heart failure [28]. Cheng et al. [29] studies have found that blood urea nitrogen levels at admission to COVID-19 patients are associated with patient mortality and can serve as a risk assessment indicator for patients with severe COVID-19. This research work also found a strong correlation between blood urea nitrogen levels and death in patients with COVID-19.

Lymphocytes are a type of white blood cell, which is an important cell of the body to play an immune response function. Studies have shown that absolute lymphocyte values are strongly associated with disease progression in patients with COVID-19, with significantly lower absolute lymphocyte values in severe patients [30]. This research work found that absolute lymphocyte values are closely related to hospitalized endotracheal intubation in patients with COVID-19.

Hypersensitive CRP is a non-specific marker of the acute phase of the inflammatory response synthesized by the liver and can be clinically used as a predictor of cardiovascular disease. Available evidence suggests that SARS-CoV-2 can damage cardiomyocytes and cause heart damage [23]. Thus, hypersensitive CRP can play a role in predicting the degree of inflammation in people with COVID-19. Guan et al. [31] studies have also found that hypersensitivity CRP can be used as a predictor of death risk in patients with COVID-19. In this research work, hypersensitivity CRP levels were strongly associated with hospital endotracheal intubation in patients with COVID-19.

Total bilirubin includes both direct bilirubin and indirect bilirubin, which are mainly used as indicators for the diagnosis of liver disease. Studies have shown that some patients with critically ill COVID-19 can develop severe liver dysfunction manifested by elevated total bilirubin, which may be associated with COVID-19 [32]. Studies by Zhan et al. [33] also found that total bilirubin levels may be a risk factor for severe COVID-19. This research work suggests that total bilirubin levels are associated with hospitalized endotracheal intubation in patients with COVID-19.

5. CONCLUSION

In summary, algorithm experts and clinicians have proposed various computer-aided diagnostic algorithms based on technologies such as image processing and machine learning to actively respond to the challenges posed by the covid-19 outbreak. This study analysed predictors with strong correlation with clinical outcomes in patients with covid-19, and extreme learning machine algorithm can better predict their in-hospital clinical outcomes, which can help guide the follow-up treatment of covid-19 patients and grasp the outcome of the disease. As next work, artificial intelligence as a data-driven discipline, collecting more high-quality data sets, can help further improve algorithm performance.

REFERENCES

[1] World Health Organization. Weekly epidemiological update on COVID-19-16 March 2021[EB/OL]. (2021-03-16) [2021-03-16]. https://www.who.int/publications/m/item/weekly-epidemiological-update--16-march-2021.

[2] Kaufmann C C, Ahmed A, Kassem M, Freynhofer M K, Jäger B, Aicher G, et al. Mid-regional pro-atrial natriuretic peptide independently predicts short-term mortality in COVID-19[J/OL]. Eur J Clin Invest, 2021, 51: e13531. DOI: 10.1111/eci.13531.

[3] Zhang R, Ouyang H, Fu L, Wang S, Han J, Huang K, et al. CT features of SARS-CoV-2 pneumonia according to clinical presentation: a retrospective analysis of 120 consecutive patients from Wuhan city[J]. Eur Radiol, 2020, 30:

4417-4426. DOI:10.1007/s00330-020-06854-1

[4] Zhang L, Yan X, Fan Q, Liu H, Liu X, Liu Z, et al. D-dimer levels on admission to predict in-hospital mortality in patients with COVID-19[J]. J ThrombHaemost, 2020, 18: 1324-1329. DOI:10.1111/jth.14859

[5] Bertsimas D, Lukin G, Mingardi L, Nohadani O, Orfanoudaki A, Stellato B, et al. COVID-19 mortality risk assessment: an international multi-center study[J/OL]. PLoS One, 2020, 15: e0243262. DOI: 10.1371/journal.pone.0243262.

[6] Zhao Z, Chen A, Hou W, Graham J M, Li H, Richman P S, et al. Prediction model and risk scores of ICU admission and mortality in COVID-19[J/OL]. PLoS One, 2020, 15: e0236618. DOI: 10.1371/journal.pone.0236618.

[7] Peng Chi, QI Ge-yao, Zhang Chen-xu, Guo Yu-feng, JinZhi-chao. Prediction of in-hospital clinical outcomes of coronavirus disease 2019 patients based on machine learning algorithms[J]. Academic Journal of Second Military Medical University, 2021, 42(10): 1115-1123.

[8] Li L, Fang X, Cheng L, Wang P, Li S, Yu H, et al. Development and validation of a prognostic nomogram for predicting in-hospital mortality of COVID-19:amulticenter retrospective cohort study of 4086 cases in China[J]. Aging (Albany NY), 2021, 13: 3176-3189.

[9] Ministry of Health and Family Welfare (GOI) - https://www.mohfw.gov.in/ (Accessed on July 2nd 2020).

[10] Chawla N V. Data mining for imbalanced datasets: an overview[M]//MAIMON O, ROKACH L. Data mining and knowledge discovery handbook. Boston, MA: Springer US, 2009: 875-886.

[11] Huang G-B, Zhu Q-Y, Siew C-K. Extreme learning machine: theory and applications. Neurocomputing 2006;70:489–501.

[12] Ojha V K, Abraham A, Snášel V. Metaheuristic design of feedforward neural networks: a review of two decades of research[J]. EngApplArtifIntell, 2017, 60: 97-116. DOI:10.1016/j.engappai.2017.01.013

[13] John G H, Langley P. Estimating continuous distributions in Bayesian classifiers[C]//Proceedings of the eleventh conference on uncertainty in artificial intelligence. Montréal, Qué, Canada: Morgan Kaufmann Publishers Inc., 1995: 338-345.

[14] Breiman L. Random forests[J]. Mach Learn, 2001, 45: 5-32. DOI:10.1023/A:1010933404324

[15] Soeters P B, Wolfe R R, Shenkin A. Hypoalbuminemia: pathogenesis and clinical significance[J]. JPEN J Parenter Enteral Nutr, 2019, 43: 181-193. DOI:10.1002/jpen.1451

[16] Li G, Zhou C L, Ba Y M, Wang Y M, Song B, Cheng X B, et al. Nutritional risk and therapy for severe and critical COVID-19 patients: a multicenter retrospective observational study[J]. Clin Nutr, 2021, 40: 2154-2161. DOI:10.1016/j.clnu.2020.09.040

[17] Violi F, Cangemi R, Romiti G F, Ceccarelli G, Oliva A, Alessandri F, et al. Is albumin predictor of mortality in COVID-19? [J]. Antioxid Redox Signal, 2021, 35: 139-142. DOI:10.1089/ars.2020.8142

[18] Horng T. Calcium signaling and mitochondrial destabilization in the triggering of the NLRP3 inflammasome[J]. Trends Immunol, 2014, 35: 253-261. DOI:10.1016/j.it.2014.02.007

[19] Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study[J]. Lancet, 2020, 395: 1054-1062. DOI:10.1016/S0140-6736(20)30566-3

[20] Zhang L K, Sun Y, Zeng H, Wang Q, Jiang X, Shang W J, et al. Calcium channel blocker amlodipine besylate therapy is associated with reduced case fatality rate of COVID-19 patients with hypertension[J/OL]. Cell Discov, 2020, 6: 96. DOI: 10.1038/s41421-020-00235-0.

[21] Neuraz A, Lerner I, Digan W, Paris N, Tsopra R, Rogier A, et al. Natural language processing for rapid response to emergent diseases: case study of calcium channel blockers and hypertension in the COVID-19 pandemic[J/OL]. J Med Internet Res, 2020, 22: e20773. DOI: 10.2196/20773.

[22] Du R H, Liang L R, Yang C Q, Wang W, Cao T Z, Li M, et al. Predictors of mortality for patients with COVID-19 pneumonia caused by SARS-CoV-2: a prospective cohort study[J/OL]. Eur Respir J, 2020, 55. DOI: 10.1183/13993003.00524-2020.

[23] Bansal M. Cardiovascular disease and COVID-19[J]. Diabetes MetabSyndr: Clin Res Rev, 2020, 14: 247-250. DOI:10.1016/j.dsx.2020.03.013

[24] Han H, Xie L L, Liu R, Yang J, Liu F, Wu K L, et al. Analysis of heart injury laboratory parameters in 273 COVID-19 patients in one hospital in Wuhan, China[J]. J Med Virol, 2020, 92: 819-823. DOI:10.1002/jmv.25809

[25] Henry B M, De Oliveira M H S, Benoit S, Plebani M, LIPPI G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019(COVID-19): a meta-analysis[J]. Clin Chem Lab Med CCLM, 2020, 58: 1021-1028. DOI:10.1515/cclm-2020-0369

[26] Li T, Wang X, Zhuang X, Wang H, Li A, Huang L, et al. Baseline characteristics and changes of biomarkers in disease course predict prognosis of patients with COVID-19[J]. Intern Emerg Med, 2021, 16: 1165-1172. DOI:10.1007/s11739-020-02560-4

[27] Kazory A. Emergence of blood urea nitrogen as a biomarker of neurohormonal activation in heart failure[J]. Am J Cardiol, 2010, 106: 694-700. DOI:10.1016/j.amjcard.2010.04.024

[28] Aronson D, Mittleman M A, Burger A J. Elevated blood urea nitrogen level as a predictor of mortality in patients admitted for decompensated heart failure[J]. Am J Med, 2004, 116: 466-473. DOI:10.1016/j.amjmed.2003.11.014

[29] Cheng A, Hu L, Wang Y, Huang L, Zhao L, Zhang C, et al. Diagnostic performance of initial blood urea nitrogen combined with D-dimer levels for predicting in-hospital mortality in COVID-19 patients[J/OL]. Int J Antimicrob Agents, 2020, 56: 106110. DOI: 10.1016/j.ijantimicag.2020.106110.

[30] Deng R, Wang C, Ye Y, Gou L, Fu Z, Ye B, et al. Clinical manifestations of blood cell parameters and inflammatory factors in 92 patients with COVID-19[J/OL]. Ann Transl Med, 2021, 9: 62. DOI: 10.21037/atm-20-7765.

[31] Guan X, Zhang B, Fu M, Li M Y, Yuan X, Zhu Y W, et al. Clinical and inflammatory features based machine learning model for fatal risk prediction of hospitalized COVID-19 patients: results from a retrospective cohort study[J]. Ann Med, 2021, 53: 257-266. DOI:10.1080/07853890.2020.1868564

[32] Fan H, Cai J, Tian A, Li Y, Yuan H, Jiang Z, et al. Comparison of liver biomarkers in 288 COVID-19 patients: a mono-centric study in the early phase of pandemic[J/OL]. Front Med (Lausanne), 2020, 7: 584888. DOI: 10.3389/fmed.2020.584888.

[33] Zhan N, Guo Y, Tian S, Huang B, Tian X, Zou J, et al. Clinical characteristics of COVID-19 complicated with pleural effusion[J/OL]. BMC Infect Dis, 2021, 21: 176. DOI: 10.1186/s12879-021-05856-8.