Article

Development of Machine-Learning Model to Predict COVID-19 Mortality: Application of Ensemble Model and Regarding Feature Impacts

Seung-Min Baik [1], Miae Lee [2], Kyung-Sook Hong [1] and Dong-Jin Park [2,*]

[1] Division of Critical Care Medicine, Department of Surgery, College of Medicine, Ewha Womans University, Seoul 07804, Korea; illaillailla@naver.com (S.-M.B.); hongks@ewha.ac.kr (K.-S.H.)

[2] Department of Laboratory Medicine, College of Medicine, Ewha Womans University of Korea, Seoul 07804, Korea; miae@ewha.ac.kr

[*] Correspondence: parkdj1280@gmail.com; Tel.: +82-2-2650-2822

Abstract: This study was designed to develop machine-learning models to predict COVID-19 mortality and identify its key features based on clinical characteristics and laboratory tests. For this, deep-learning (DL) and machine-learning (ML) models were developed using receiver operating characteristic (ROC) area under the curve (AUC) and F1 score optimization of 87 parameters. Of the two, the DL model exhibited better performance (AUC 0.8721, accuracy 0.84, and F1 score 0.76). However, we also blended DL with ML, and the ensemble model performed the best (AUC 0.8811, accuracy 0.85, and F1 score 0.77). The DL model is generally unable to extract feature importance; however, we succeeded by using the Shapley Additive exPlanations method for each model. This study demonstrated both the applicability of DL and ML models for classifying COVID-19 mortality using hospital-structured data and that the ensemble model had the best predictive ability.

Keywords: COVID-19; mortality; artificial intelligence; ensemble model

1. Introduction

In December 2019, the World Health Organization (WHO) first reported pneumonia of an unknown cause, which was subsequently linked to a novel coronavirus [1]. In February 2020, the WHO officially identified severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as the causative agent of coronavirus disease 2019 (COVID-19), which spread rapidly, prompting the WHO to declare a worldwide pandemic in March 2020. Despite the passage of more than two years since the outbreak, the pandemic persists, notably in Korea, where the number of COVID-19-related deaths recently increased sharply [2], and several trials have been conducted in various fields, including treatment [3,4].

Most cases of COVID-19 are asymptomatic or only involve mild symptoms [5–10]. Nevertheless, the number of critically ill patients and fatalities continues to increase worldwide. In January 2022, the WHO reported that more than 370 million people had developed COVID-19, and more than 5.6 million had died [11]. The U.S. Centers for Disease Control and Prevention reported that, among all COVID-19 patients in the United States, 14% required hospitalization, 2% required intensive care, and 5% died [12]. High mortality rates were reported in areas with a high incidence of COVID-19 [13,14]. An important contributing factor is the inability of limited medical resources to manage the rapidly increasing number of patients, which results in the failure to diagnose COVID-19 in a timely manner, thereby contributing to increased mortality.

The Sequential Organ Failure Assessment, Acute Physiology and Chronic Health Evaluation II, and Simplified Acute Physiology scores are commonly used as scales for mortality prediction in critically ill patients [15–17]. However, they have limitations in the risk stratification necessary for mortality prediction [18]. Recently, analyses of mortality

prediction via machine learning (ML) and deep learning (DL) have been conducted in clinical settings. Multiple studies using ML demonstrated the successful prediction of COVID-19 severity and mortality [19–21]. An ML study predicted the deterioration of a patient's condition using demographic and comorbidity data and vital signs [22]. In addition, several DL models were published [23,24]. However, there have been few ML or DL studies of COVID-19 conducted in Korea.

ML facilitates the resolution of unknown patterns in data by learning training data and making models to achieve better predictions [25,26]. Several ML algorithms (e.g., support vector machine (SVM), random forest (RF), and boosting models) were used in biology, genomics, and toxicology [27–31]. In medical fields, with the exception of image and text data in medical charts, most laboratory tests use data from blood or urine samples, which are mainly composed of numerical inputs [32]. A deep neural network (DNN), a type of DL that uses hidden layers [33,34], is renowned for its ability to analyze high-dimensional data. Because bioinformatic data are generally high-dimensional, a DNN may be a suitable model for the study of bioinformatics [35,36].

There are several studies on the usefulness of DL using image data from COVID-19 patients [37,38]; however, studies using numerical data, which account for most medical data, are still insufficient.

This study was performed to analyze the electronic medical record data and laboratory findings of hospitalized COVID-19 patients in Korea using ML and DL to develop an optimized model for mortality prediction.

2. Materials and Methods

The principal procedures of the machine-learning prediction model for COVID-19 mortality are described in the following sections. The overall workflow is demonstrated in Figure 1.
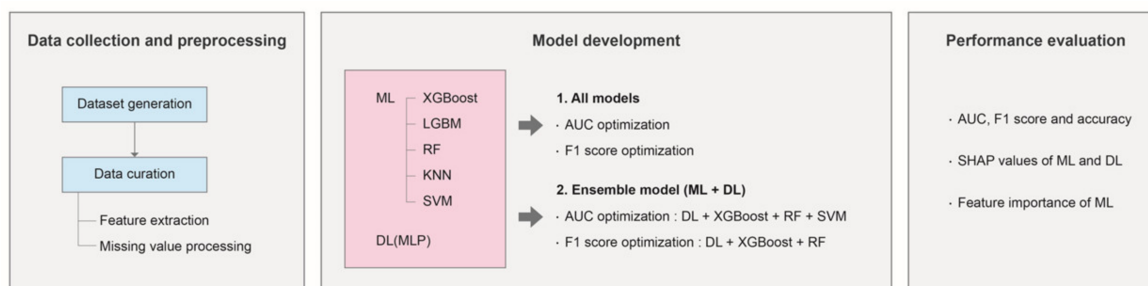
Figure 1. Schematic overall flow chart of conducting the study.

2.1. Patients and Data Collection

This study included patients diagnosed with COVID-19 who were admitted to the sub-intensive and intensive care units between September 2021 and January 2022. All had moderate-to-severe COVID-19. Moderate severity was defined as requiring oxygen supplied by a high-flow system or mechanical ventilation. Data were collected regarding the following clinical characteristics: sex, age, medical history, vital signs, chief complaints, review of systems, and mortality.

Data were also collected regarding the following laboratory data: complete blood count with differential count, prothrombin time, activated partial thromboplastin time, total calcium, phosphorus, glucose, blood urea nitrogen (BUN), creatinine, estimated glomerular filtration rate, triglyceride, total cholesterol, total protein, albumin, aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase, total bilirubin, sodium, potassium, chloride, total carbon dioxide (CO_2), magnesium, amylase, lipase, C-reactive protein, ammonia, arterial blood gas analysis, lactate, creatinine kinase (CK), CK-MB, high sensitivity troponin T, procalcitonin, lactate dehydrogenase, N-terminal pro-B-type natri-

uretic peptide, uric acid, serum osmolarity, ferritin, fibrinogen, fibrinogen degradation production, and D-dimer.

All clinical characteristics and laboratory findings were the initial values obtained after admission. This study was approved by the Institutional Review Board (IRB) of Ewha Womans University Mokdong Hospital (approval number: EUMC 2022-01-031), which waived the need for informed consent. All methods and datasets were carried out in accordance with relevant guidelines and regulations by the IRB.

2.2. Statistical Data Analyses

COVID-19 mortality was evaluated using the chi-squared test and the Student's *t*-test. All statistical analyses were performed using IBM SPSS software, version 26.0 (SPSS Inc., Chicago, IL, USA).

2.3. DL Model Selection and Training

The DL model was shown to approximate the function of a complex structure [39–41]. Ours was conducted using a multi-layer perceptron (MLP), a class of DNN, for classification [42,43]. An MLP consists of at least three layers: input, hidden, and output. In this study, the hidden layer comprised one layer to which we applied the dropout technique, a simple method to prevent overfitting in neural networks [44–46]. Our MLP model was updated using the Adam optimizer, a method of stochastic optimization that uses a calculated parameter gradient.

2.4. ML Model Selection

Several ML models were used for disease prediction. Among the decision-tree models, the bagging and boosting ML models demonstrated a robust classification of COVID-19 mortality. In addition, SVM and K-nearest neighbors (KNN) models were used to create an optimized model for mortality classification. In this study, there were 8.0% missing values in 17,052 datasets from 203 COVID-19 cases.

2.5. ML Model Development

Tree-based models can be implemented regardless of missing values. However, SVM and KNN models cannot; instead, the missing values are replaced with the median values of each test result. Here, we used the following ML models: RF, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LGBM), SVM, and KNN.

2.6. Development of an Ensemble Model Using DL and ML

An ensemble model was developed with the best-performing ML models and a DL (MLP) model to achieve optimized results. First, the area under the receiver operating characteristic curve (AUC) optimization was conducted by integrating the results of four models: three ML models (XGBoost, RF, and SVM) and one DL model (MLP). Second, for F1 score optimization, an ensemble model was developed by hard voting that involved the top-three F1 score models: XGBoost, RF, and DL. The ensemble model was developed by using only the results identified as COVID-19 mortality in ≥2 models to determine the optimized F1 score.

2.7. Performance Measurement

We used the F1 score, accuracy, and AUC data as performance measures for the model, which was evaluated using the AUC analysis provided by Python Library.

2.8. Shapley Additive exPlanations (SHAP) Method

The SHAP method is a new technique for estimating the influence of each feature according to a probabilistic game rule [47,48]. In DL models, it is impossible to determine the extent to which each parameter (feature) contributes to completed model performance because DL models are effectively "black box" systems. In order to solve this problem, we

used the SHAP method, which was also used to predict the feature impacts of three ML models: XGBoost, LGBM, and RF.

2.9. Feature Extraction

Feature extraction is important for developing and measuring the performances of ML models, and in our study, we identified three new parameter ratios: pulse rate (PR)/respiratory rate (RR), AST/ALT, and neutrophil/lymphocyte.

2.10. Standard Scaling

As a data pre-processing step during ML and DL model training, the range of columns in the COVID-19 dataset was standardized and scaled using "scikit-learn" in the Python Library.

2.11. Cross-Validation

For K-fold cross-validation, we divided the total dataset of 203 COVID-19 patients to create the training and validation sets using a ratio of 4:1, respectively. Because we had a limited number of cases, the test set itself may not have fully represented all possible clinical characteristics and laboratory findings. Thus, K-fold validation (n_split:5) was applied to the entire dataset to overcome this limitation.

3. Results

3.1. Demographic and Clinical Characteristics

This study included 203 Korean patients between September 2021 and January 2022 who had moderate-to-severe COVID-19. Of these, 49 (23.1%) were in the non-survivor group, and 154 were in the survivor group (76.9%). The mean age was significantly older in the non-survivor group ($p < 0.05$). The initial demographics and clinical characteristics of the two groups are shown in Table 1.

Table 1. Demographics and clinical characteristics of COVID-19 patients.

| Variable	Total (n = 203)	Non-Survival (n = 49)	Survival (n = 154)	p
Male sex (%)	98 (48.3)	24 (49.0)	74 (48.1)	0.910
Age (years)	67.5 ± 16.16	75.7 ± 11.58	64.9 ± 16.57	<0.05
Hospitalization period	13.4 ± 10.20	16.8 ± 12.81	12.4 ± 9.00	<0.05
Comorbidities (%)				
Hypertension	98 (48.3)	30 (61.2)	68 (44.2)	0.055
Diabetes mellitus	56 (27.6)	20 (40.8)	36 (23.4)	<0.05
Heart disease	6 (3.0)	6 (12.2)	10 (6.5)	0.319
Lung diasease	12 (5.9)	3 (6.1)	9 (5.8)	0.943
Liver disease	5 (2.5)	0 (0)	5 (3.2)	0.454
Kidney disease	5 (2.5)	2 (4.1)	3 (1.9)	0.756
Brain disease	27 (13.3)	6 (12.2)	21 (13.6)	0.993
Malignant disease	28 (13.8)	7 (14.3)	21 (13.6)	0.909
Vital signs at hospital admission				
Systolic blood pressure (mmHg)	133.4 ± 22.08	132.9 ± 25.12	133.6 ± 21.10	0.866
Diastolic blood pressure (mmHg)	79.3 ± 14.76	75.2 ± 14.22	80.6 ± 14.74	<0.05
Pulse rate (PR, bpm)	84.4 ± 17.79	84.3 ± 23.84	84.5 ± 15.53	0.962
Respiratory rate (RR, bpm)	22.4 ± 10.03	25.6 ± 17.10	21.3 ± 6.17	0.095
Body temperature (°C)	37.1 ± 0.71	36.9 ± 0.82	37.1 ± 0.65	<0.05
Pulse pressure (mmHg)	54.1 ± 17.88	57.7 ± 20.23	53.0 ± 16.99	0.147
PR/RR	4.1 ± 1.34	3.9 ± 1.74	4.2 ± 1.18	0.316

3.2. Data Features and Parameters

For ML model evaluation, because of the data imbalance between the non-survivor and survivor groups, we evaluated the performances of individual models using the AUC and F1 scores.

In our study, 84 data parameters were collected for each patient: 65 laboratory test results and 19 that included medical history obtained from the patients' electronic medical records, blood pressure, body temperature, and arterial blood gas analysis (Table 1, Supplementary Table S1). Three new parameters were obtained by feature extraction: the AST/ALT, neutrophil/lymphocyte, and PR/RR ratios, which were used to develop the ML models, including the MLP of the DL model. Thus, 87 mostly numerical and structured parameters were used for model development, with the goal of creating a classifier ML model and a neural network (i.e., a DL model) to distinguish between non-survivor and survivor groups.

3.3. DL (Neural Network) Performance Optimized by the AUC and F1 Score

We developed a new DL neural network and trained it using the 87 parameters. AUC optimization revealed a value of 0.8721, an accuracy of 0.84, an F1 score of 0.76, a precision of 0.79, and a recall of 0.74 (Figure 2 and Table 2). The DL model exhibited the best accuracy, F1 score, and recall compared to the ML models; thus, DL performed better than ML for classifying numerical and structured patient data.
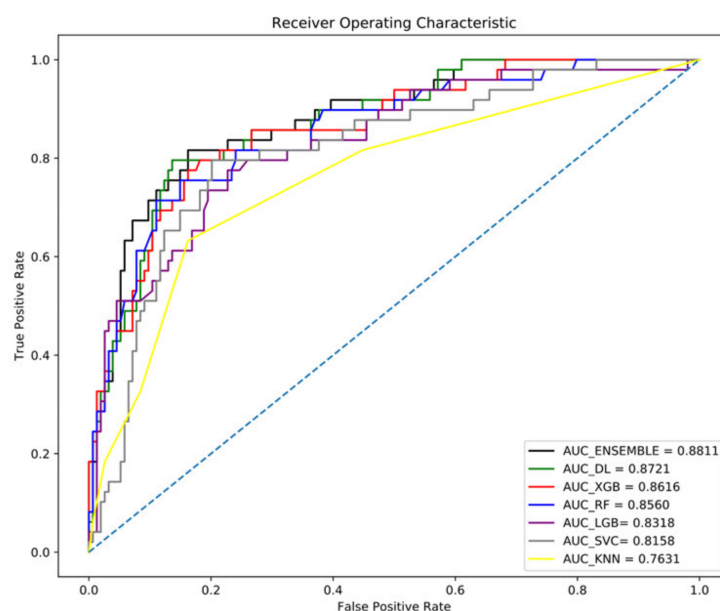
Figure 2. Area under the receiver operating characteristic curves (AUC) of machine-learning models, deep-learning model, and an ensemble model.

Table 2. DL and ML performances by AUC optimization.

Classifier	AUC	Accuracy	F1-Score	Precision	Recall
XGboost	0.8616	0.82	0.75	0.8	0.73
LGBM	0.8318	0.83	0.71	0.83	0.68
RF	0.8560	0.83	0.74	0.80	0.71
KNN	0.7631	0.79	0.59	0.77	0.58
SVM	0.8158	0.81	0.72	0.74	0.71
DL	0.8721	0.84	0.76	0.79	0.74
Ensemble model *	0.8811	0.85	0.77	0.81	0.75

DL: deep learning; ML: machine learning; XGboost: extreme gradient boosting; LGBM: light gradient boosting model; RF: random forest; KNN: K-nearest neighbors; SVM: support vector machine. * Ensemble model of deep-learning and machine-learning models.

In addition, optimized cut-offs were determined for each ML model, and hyperparameter tuning was used to achieve an optimized F1 score. The optimized DL model had an F1 score of 0.78, an AUC of 0.8614, an accuracy of 0.83, a precision of 0.77, and a recall of 0.79

(Table 3). The F1 score for the DL optimization model was 0.78, which was the same as RF, the highest result (Table 3). The accuracy of our DL model was 0.83, and its results were similar to the findings of the tree-based models (RF, XGBoost, and LGBM).

Table 3. DL and ML performances by F1 score optimization.

Classifier	AUC	Accuracy	F1-Score	Precision	Recall
XGboost	0.8331	0.83	0.77	0.76	0.77
LGBM	0.8318	0.85	0.75	0.84	0.72
RF	0.8560	0.84	0.78	0.78	0.77
KNN	0.7631	0.79	0.72	0.72	0.74
SVM	0.8158	0.82	0.76	0.76	0.76
DL	0.8614	0.83	0.78	0.77	0.79
Ensemble model *	0.8631	0.85	0.80	0.80	0.80

DL: deep learning; ML: machine learning; XGboost: extreme gradient boosting; LGBM: light gradient boosting model; RF: random forest; KNN: K-nearest neighbors; SVM: support vector machine. * Ensemble model of deep-learning and machine-learning models.

3.4. Performances of ML Models Optimized by AUC

Tree-series models generally demonstrate good performance in solving classification problems [49–51]. Therefore, we developed ML models that used tree-based models (RF, XGBoost, and LGBM) and other ML approaches (SVM and KNN). The performance results for each model are described in Table 2. The best performance among the ML models was demonstrated by XGBoost (AUC: 0.8616), a boosting series of tree-based models; the next best performances were by RF (AUC: 0.8560), the boosting model LGBM (AUC: 0.8318), SVM (AUC: 0.8158, based on the relative distance between data), and the KNN model (AUC: 0.7631) (Figure 2). The RF and LGBM models demonstrated the highest accuracy (0.83), while XGBoost (0.82), SVM (0.81), and KNN (0.79) were slightly lower (Table 2).

3.5. ML Model Results Optimized According to F1 Score

In this study, data imbalance was present between the non-survivor and survivor groups. Thus, the F1 score was used as the evaluation index for model performance evaluation and optimization (Table 3). Similar to the AUC results, the DL model showed the best F1 score. Among the ML models, RF had the highest F1 score (0.78) followed by XGBoost (0.77), SVM (0.76), LGBM (0.75), and KNN (0.72) (Table 3).

### 3.6. Ensemble Model of DL and ML Models

Because DL exhibited good performance in AUC analyses, we developed a new ensemble model that blended the DL model with three ML models based on their AUC outcomes. The greatest weight was placed on the DL model, followed by XGBoost, RF, and SVM. The highest-performing ensemble model achieved the best AUC result (0.8811) and demonstrated the highest accuracy (0.85) (Figure 2 and Table 2). Finally, we created an optimal AUC ensemble model that showed 85% accuracy, 81% precision, 75% recall, and an F1 score of 77% (Table 2).

Additionally, to achieve F1 score optimization through the ensemble, we selected three models—RF, XGBoost, and the DL model with the highest F1 score. We used the hard voting technique to vote in two or more models. The resulting highest F1 score was 0.80 (Table 3). The DL model showed a better AUC and F1 score than the ML model. As a result, it had better performance (Figure 2 and Tables 2 and 3).

3.7. Feature Importances in ML Models

The top 20 important features in the RF and XGBoost models are shown in Figure 3. In XGBoost, BUN was the most important; the others were base excess, pH, saturation of percutaneous oxygen (SpO$_2$), partial pressure of oxygen (PO$_2$), and procalcitonin. In the

RF model, pH was the most important feature, followed by BUN, SpO$_2$, base excess, PO$_2$, and total calcium.


(a)

(b)
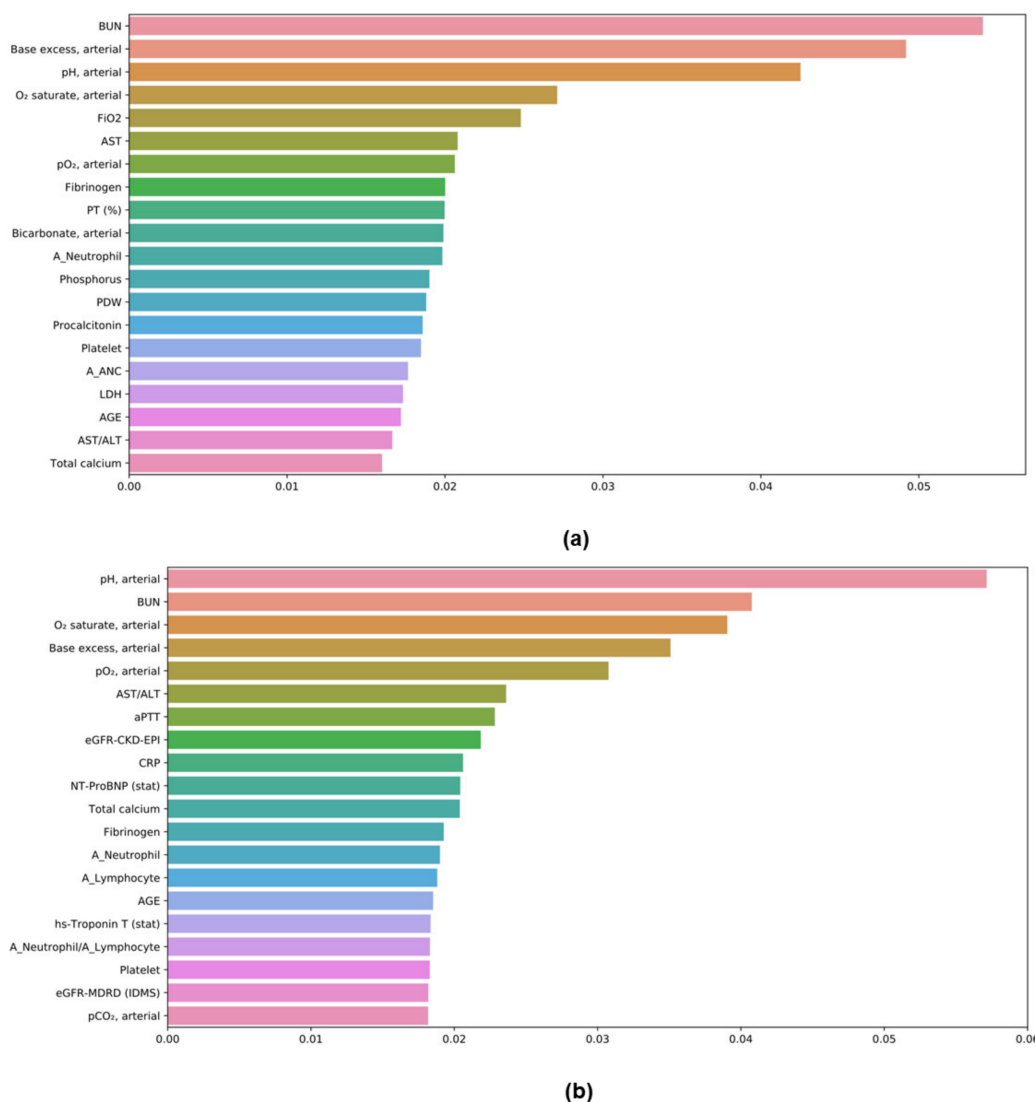
Figure 3. Feature importance of XGBoost (**a**) and RF (**b**) in COVID-19 mortality.

3.8. Evaluation of the Model Performance Impact When Using the Mean SHAP Value in DL and ML Models

The SHAP method provided the means for assessing the effects of features on classification targets. We analyzed the mean SHAP values of the DL and three ML models to identify features associated with COVID-19 mortality. The top 20 features that affected each model are shown in Figure 4. In DL and RF, pH was the most important feature; it was negatively associated with disease outcomes in the non-survivor group. In LGBM and XGBoost, partial pressure of carbon dioxide (PCO$_2$) was the most important feature and was positively associated with disease outcomes in the non-survivor group. The newly identified variable, the AST/ALT ratio, was included in the top 20 features of DL, RF, XGBoost, and LGBM models.

(a)

(b)

(c)

Figure 4. *Cont.*

(d)
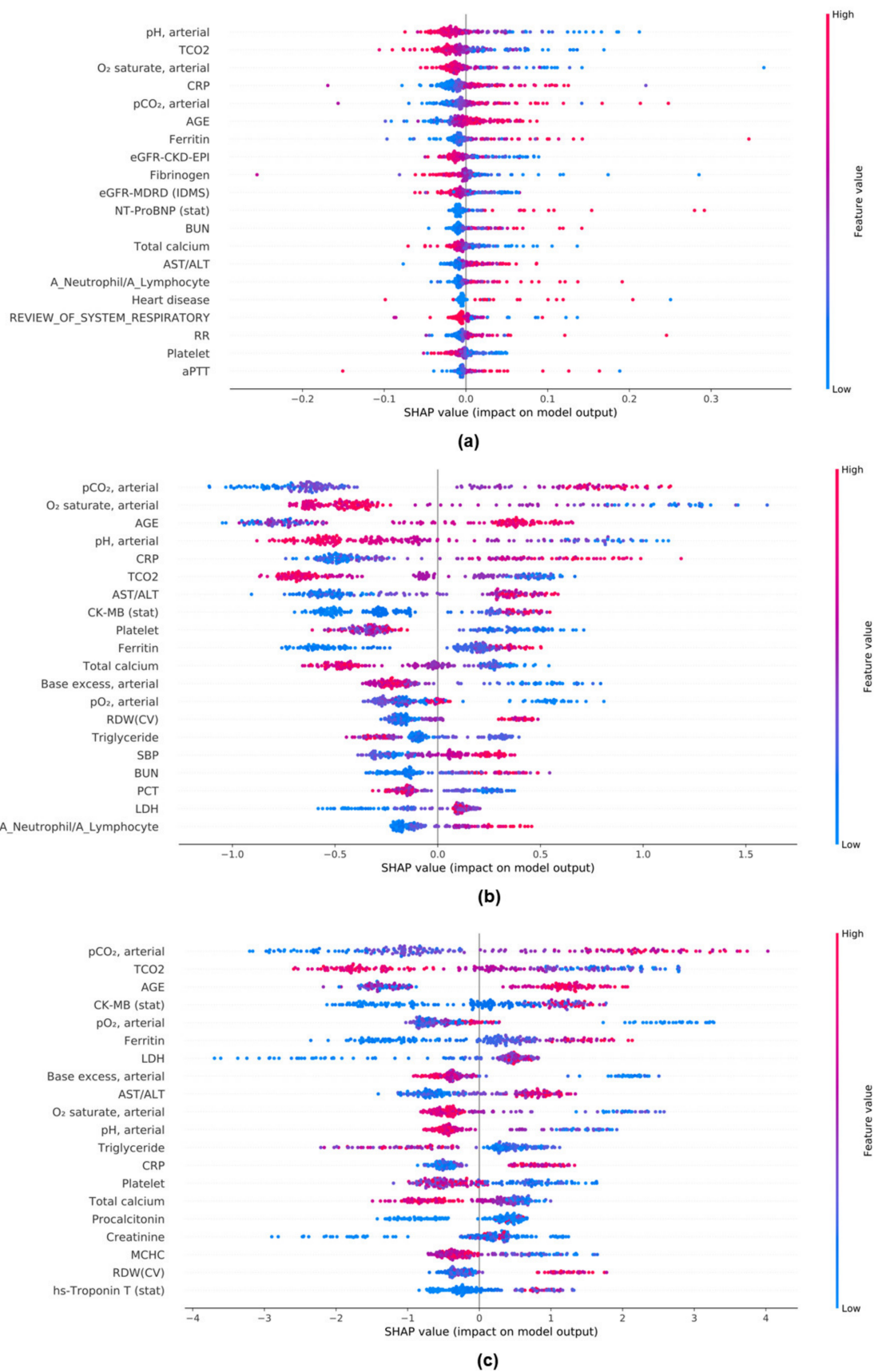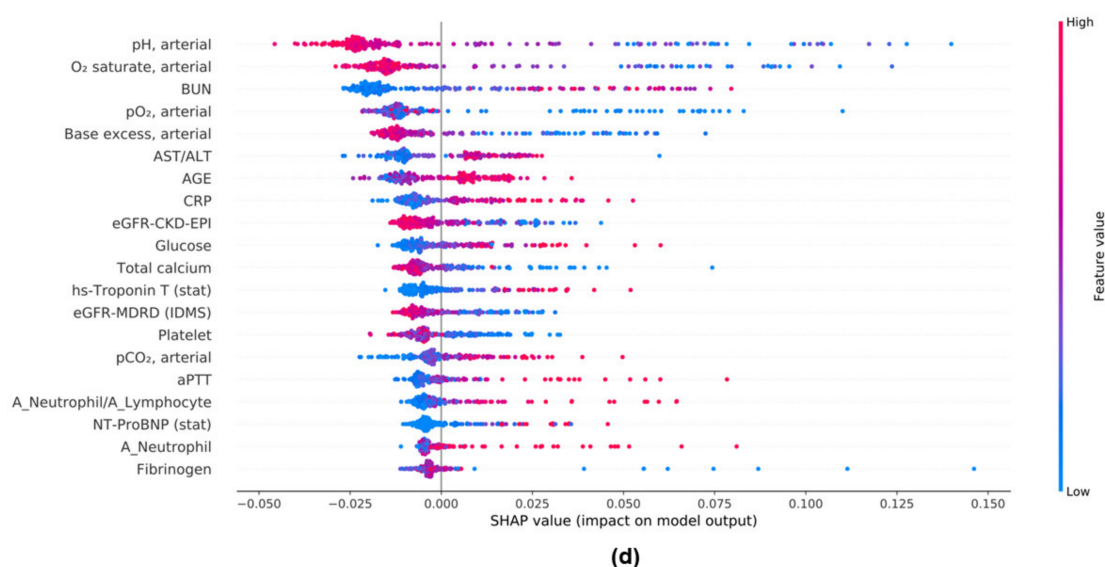
Figure 4. Analysis of features contributing to COVID-19 mortality by the SHAP method. (**a**) DL model. (**b**) XGBoost model. (**c**) LGBM model. (**d**) RF model.

4. Discussion

In this study, we used 87 parameters, including laboratory tests and clinical characteristics, to develop an ensemble model from DL and ML models to predict COVID-19 mortality. Although the study did not include a large number of patients, a large and varied parameter set was used to develop predictive artificial intelligence models; additionally, the model was trained by performing K-fold validation to avoid loss in the test set. In this study, a DL model and several ML models were evaluated using AUC analysis and the F1 score. The DL model showed the best performance for accuracy and F1 score compared to the ML models (XGBoost, LGBM, RF, SVM, and KNN, Tables 2 and 3).

Several studies used ML or DL alone to predict COVID-19 mortality, such as an ML study by Vaid A et al., to predict mortality in 4098 patients [52]. According to reports, XGBoost was the best, with an AUC of 0.84 to 0.90, which was similar to our results (Figure 2). Li, Xiaoran et al., reported a mortality prediction model using deep learning for 1108 patients, of whom 142 died [53]. The AUC of the deep-learning model in this study was 0.844, which was lower than that of our DL and ensemble models (Figure 2). Although our study was conducted with a relatively small number of cases, the model performances were similar to those of other studies. There are two possible reasons.

First, the tree-series models generally demonstrated good performance with categorical structured data. However, the DL model performed better than the ML models because the data from our study consisted mostly of laboratory test results, which comprise numerical data. Categorical parameters, including sex and medical history (e.g., diabetes, hypertension, and heart disease), were converted into binary data using a label-encoding library. Therefore, compared to DL, the potential gain from the ML tree models was greatly reduced. Specifically, the shallow MLP of the DL model was effective for solving the classification problem because most components of hospital data are numerical. This result is consistent with the findings of other studies that used numerical hospital data [32].

The second reason is that the number of columns (parameters) was comparatively higher than the number of rows (number of patient cases) in our data structure. When there are comparatively more parameters than cases, a tree model is more likely to demonstrate poor performance because of overfitting. Because the number of cases was small and the number of columns large, there was an increasing number of questions during tree creation, so the tree became deeper. Therefore, dimensionality may impede model learning or cause overfitting. The DL model of the MLP was a combination of multiple layers of linear models. Thus, the number of columns was much less important in the linear model

than in the tree model. Accordingly, the MLP showed good performance. However, the best performance was obtained when one hidden layer of the MLP was used. When two or more hidden layers were stacked, the performance was poor because of overfitting.

The dropout technique was notably used in this study. Without it, our model would have had to learn mainly through the features (512 nodes) that are considered important in the training set. This can lead to overfitting, resulting in poor outcomes in the test set. The advantage of the dropout method is that the model learns from fewer features (nodes) that are considered important but also from features (nodes) that are presumably negligible; thus, it has generalized power to provide a good fit even if unusual data subsequently enter the test set.

We developed ensemble machine-learning models to predict COVID-19 mortality using a DL model combined with several ML models. We found that the ensemble model of DL and ML achieved the optimal classification results in hospital-structured data (Tables 2 and 3).

DL models generally cannot extract feature importance because they use a "black-box" approach; however, we did so through the SHAP method for each model, including the DL model. For data analysis and interpretation, the SHAP method was better than the widely used feature importance technique because it only lists the feature importance order for each model; it cannot accurately visualize whether each feature has a positive or negative influence on the classification target. In contrast, the SHAP method can indicate the degree and direction of each parameter's effects on the classification target. In the DL model with the best performance, pH had the highest impact; in the non-survivor group, pH demonstrated a negative association with disease outcomes because of its relationship with acidosis, which is strongly associated with patient mortality [54]. Total CO_2 also had a large impact. Because CO_2 exists in the body as bicarbonate, analyzing total CO_2 provides information on the patient's acid–base balance; a lower total CO_2 value is associated with a greater likelihood of acidosis [55]. Total CO_2 also demonstrated a negative association in the non-survivor group, which suggested acidosis. Although the Student's *t*-test found that the PCO_2 level did not differ significantly between the two groups, SHAP revealed a positive association and high impact in the non-survivor group. High PCO_2 levels are associated with respiratory acidosis and may lead to mortality. This study was conducted by generating new variables from existing laboratory findings. Among them, the neutrophil/lymphocyte ratio and the AST/ALT ratio showed a positive association with disease outcomes in the non-survivor group.

As is the case with other COVID-19 patients [56], a decrease in lymphocytes and an increase in neutrophils were observed in our study, but the levels between the two groups differed significantly; however, these variables demonstrated no feature effect in the DL model, as determined using the SHAP method. However, the neutrophil/lymphocyte ratio had a large impact. In addition, feature analysis in XGBoost, LGBM, and RF models was performed using the SHAP method. In the RF, XGBoost, and DL models, BUN had a high impact and showed a positive association with disease outcomes in the non-survivor group (Figure 4a,b,d), but high BUN levels also indicated decreased kidney function. COVID-19 causes sepsis, which causes dysfunction in various organs, including the kidneys. In patients with sepsis, fluid demand may increase [57], and BUN rises in the absence of an adequate fluid supply. Indeed, BUN was significantly higher in the non-survivor group. In the XGBoost, LGBM, RF, and DL models, the AST/ALT ratio showed a positive association with disease outcomes in the non-survivor group. Elevated AST and ALT were also observed [56], and in our study, these levels tended to be higher in the non-survivor group, but the results were not statistically significant. However, the AST/ALT ratio was significantly higher in the non-survivor group; its feature impact was high in the XGBoost, LGBM, RF, and DL models. The importance of AST and ALT in respiratory diseases may be low in clinical practice; however, we found that they were important indicators for COVID-19 mortality. In contrast to tree-based ML models (e.g., XGBoost, LGBM, and RF),

only the DL model demonstrated how features affect heart disease and a patient's general condition. Our findings confirmed that capabilities differed between DL and ML models.

In summary, this study demonstrated the application of DL (MLP) and ML in classifying COVID-19 mortality using numerical data (e.g., laboratory test results and clinical characteristics) from hospitalized patients. Although the performances of the ML models were relatively good, the DL model performed the best overall. In addition, using the DL and ML ensemble model, we achieved the most optimized performance. Finally, our approach identified several novel parameters for predicting COVID-19 mortality.

In conclusion, we demonstrated that DL and ML obtain favorable outcomes for mortality prediction using hospital data. Our ML models shed new light on the values of clinical information and laboratory tests that had not been previously highlighted.

References

1. World Health Organization. Coronavirus Disease (COVID-19). Available online: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed on 31 December 2019).
2. Korea Disease Control and Prevention Agency. Available online: https://www.kdca.go.kr/ (accessed on 15 February 2022).
3. Nedyalkova, M.; Vasighi, M.; Sappati, S.; Kumar, A.; Madurga, S.; Simeonov, V. Inhibition Ability of Natural Compounds on Receptor-Binding Domain of SARS-CoV2: An In Silico Approach. *Pharmaceuticals* **2021**, *14*, 1328. [CrossRef] [PubMed]
4. Nedyalkova, M.; Simeonov, V. Partitioning Pattern of Natural Products Based on Molecular Properties Descriptors Representing Drug-Likeness. *Symmetry* **2021**, *13*, 546. [CrossRef]
5. Sutton, D.; Fuchs, K.; D'Alton, M.; Goffman, D. Universal Screening for SARS-CoV-2 in Women Admitted for Delivery. *N. Engl. J. Med.* **2020**, *382*, 2163–2164. [CrossRef] [PubMed]
6. Baggett, T.P.; Keyes, H.; Sporn, N.; Gaeta, J.M. Prevalence of SARS-CoV-2 Infection in Residents of a Large Homeless Shelter in Boston. *JAMA* **2020**, *323*, 2191–2192. [CrossRef] [PubMed]
7. Campbell, K.H.; Tornatore, J.M.; Lawrence, K.E.; Illuzzi, J.L.; Sussman, L.S.; Lipkind, H.S.; Pettker, C.M. Prevalence of SARS-CoV-2 Among Patients Admitted for Childbirth in Southern Connecticut. *JAMA* **2020**, *323*, 2520–2522. [CrossRef] [PubMed]
8. Louie, J.K.; Scott, H.M.; DuBois, A.; Sturtz, N.; Lu, W.; Stoltey, J.; Masinde, G.; Cohen, S.; Sachdev, D.; Philip, S.; et al. Lessons from Mass-Testing for Coronavirus Disease 2019 in Long-Term Care Facilities for the Elderly in San Francisco. *Clin. Infect. Dis.* **2021**, *72*, 2018–2020. [CrossRef]
9. Kasper, M.R.; Geibe, J.R.; Sears, C.L.; Riegodedios, A.J.; Luse, T.; Von Thun, A.M.; McGinnis, M.B.; Olson, N.; Houskamp, D.; Fenequito, R.; et al. An Outbreak of Covid-19 on an Aircraft Carrier. *N. Engl. J. Med.* **2020**, *383*, 2417–2426. [CrossRef]
10. Letizia, A.G.; Ramos, I.; Obla, A.; Goforth, C.; Weir, D.L.; Ge, Y.; Bamman, M.M.; Dutta, J.; Ellis, E.; Estrella, L.; et al. SARS-CoV-2 Transmission among Marine Recruits during Quarantine. *N. Engl. J. Med.* **2020**, *383*, 2407–2416. [CrossRef]
11. World Health Organization. Dashboard. Available online: https://covid19.who.int/ (accessed on 20 January 2022).
12. Stokes, E.K.; Zambrano, L.D.; Anderson, K.N.; Marder, E.P.; Raz, K.M.; El Burai Felix, S.; Tie, Y.; Fullerton, K.E. Coronavirus Disease 2019 Case Surveillance—United States, January 22–May 30, 2020. *MMWR Morb. Mortal. Wkly. Rep.* **2020**, *69*, 759–765. [CrossRef]

13. Zhang, Z.; Yao, W.; Wang, Y.; Long, C.; Fu, X. Wuhan and Hubei COVID-19 mortality analysis reveals the critical role of timely supply of medical resources. *J. Infect.* **2020**, *81*, 147–178. [CrossRef]

14. Shim, E.; Mizumoto, K.; Choi, W.; Chowell, G. Estimating the Risk of COVID-19 Death During the Course of the Outbreak in Korea, February–May 2020. *J. Clin. Med.* **2020**, *9*, 1641. [CrossRef] [PubMed]

15. Knaus, W.A.; Draper, E.A.; Wagner, D.P.; Zimmerman, J.E. APACHE II: A severity of disease classification system. *Crit. Care Med.* **1985**, *13*, 818–829. [CrossRef] [PubMed]

16. Le Gall, J.R.; Lemeshow, S.; Saulnier, F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* **1993**, *270*, 2957–2963. [CrossRef] [PubMed]

17. Vincent, J.L.; Moreno, R.; Takala, J.; Willatts, S.; De Mendonca, A.; Bruining, H.; Reinhart, C.K.; Suter, P.M.; Thijs, L.G. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med.* **1996**, *22*, 707–710. [CrossRef]

18. Kądziołka, I.; Świstek, R.; Borowska, K.; Tyszecki, P.; Serednicki, W. Validation of APACHE II and SAPS II scales at the intensive care unit along with assessment of SOFA scale at the admission as an isolated risk of death predictor. *Anaesthesiol. Intensive Ther.* **2019**, *51*, 107–111. [CrossRef]

19. Mahdavi, M.; Choubdar, H.; Zabeh, E.; Rieder, M.; Safavi-Naeini, S.; Jobbagy, Z.; Ghorbani, A.; Abedini, A.; Kiani, A.; Khanlarzadeh, V.; et al. A machine learning based exploration of COVID-19 mortality risk. *PLoS ONE* **2021**, *16*, e0252384. [CrossRef]

20. Kim, H.J.; Han, D.; Kim, J.H.; Kim, D.; Ha, B.; Seog, W.; Lee, Y.K.; Lim, D.; Hong, S.O.; Park, M.J.; et al. An Easy-to-Use Machine Learning Model to Predict the Prognosis of Patients With COVID-19: Retrospective Cohort Study. *J. Med. Internet Res.* **2020**, *22*, e24225. [CrossRef]

21. Jakob, C.E.M.; Mahajan, U.M.; Oswald, M.; Stecher, M.; Schons, M.; Mayerle, J.; Rieg, S.; Pletz, M.; Merle, U.; Wille, K.; et al. Prediction of COVID-19 deterioration in high-risk patients at diagnosis: An early warning score for advanced COVID-19 developed by machine learning. *Infection* **2022**, *50*, 359–370. [CrossRef]

22. Singhal, L.; Garg, Y.; Yang, P.; Tabaie, A.; Wong, A.I.; Mohammed, A.; Chinthala, L.; Kadaria, D.; Sodhi, A.; Holder, A.L.; et al. eARDS: A multi-center validation of an interpretable machine learning algorithm of early onset Acute Respiratory Distress Syndrome (ARDS) among critically ill adults with COVID-19. *PLoS ONE* **2021**, *16*, e0257056. [CrossRef]

23. Liang, W.; Yao, J.; Chen, A.; Lv, Q.; Zanin, M.; Liu, J.; Wong, S.; Li, Y.; Lu, J.; Liang, H.; et al. Early triage of critically ill COVID-19 patients using deep learning. *Nat. Commun.* **2020**, *11*, 3543. [CrossRef]

24. Alafif, T.; Tehame, A.M.; Bajaba, S.; Barnawi, A.; Zia, S. Machine and Deep Learning towards COVID-19 Diagnosis and Treatment: Survey, Challenges, and Future Directions. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1117. [CrossRef] [PubMed]

25. Mortazavi, B.J.; Bucholz, E.M.; Desai, N.R.; Huang, C.; Curtis, J.P.; Masoudi, F.A.; Shaw, R.E.; Negahban, S.N.; Krumholz, H.M. Comparison of Machine Learning Methods with National Cardiovascular Data Registry Models for Prediction of Risk of Bleeding After Percutaneous Coronary Intervention. *JAMA Netw. Open* **2019**, *2*, e196835. [CrossRef] [PubMed]

26. Wainberg, M.; Merico, D.; Delong, A.; Frey, B.J. Deep learning in biomedicine. *Nat. Biotechnol.* **2018**, *36*, 829–838. [CrossRef] [PubMed]

27. Zhang, Z. Introduction to machine learning: K-nearest neighbors. *Ann. Transl. Med.* **2016**, *4*, 218. [CrossRef]

28. Zhang, Y.; Xie, R.; Wang, J.; Leier, A.; Marquez-Lago, T.T.; Akutsu, T.; Webb, G.I.; Chou, K.C.; Song, J. Computational analysis and prediction of lysine malonylation sites by exploiting informative features in an integrative machine-learning framework. *Brief. Bioinform.* **2019**, *20*, 2185–2199. [CrossRef]

29. Yaman, E.; Subasi, A. Comparison of Bagging and Boosting Ensemble Machine Learning Methods for Automated EMG Signal Classification. *BioMed Res. Int.* **2019**, *2019*, 9152506. [CrossRef]

30. Sarica, A.; Cerasa, A.; Quattrone, A. Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. *Front. Aging Neurosci.* **2017**, *9*, 329. [CrossRef]

31. Maltarollo, V.G.; Gertrudes, J.C.; Oliveira, P.R.; Honorio, K.M. Applying machine learning techniques for ADME-Tox prediction: A review. *Expert Opin. Drug Metab. Toxicol.* **2015**, *11*, 259–271. [CrossRef]

32. Park, D.J.; Park, M.W.; Lee, H.; Kim, Y.J.; Kim, Y.; Park, Y.H. Development of machine learning model for diagnostic disease prediction based on laboratory tests. *Sci. Rep.* **2021**, *11*, 7567. [CrossRef]

33. Creatore, C.; Sabathiel, S.; Solstad, T. Learning exact enumeration and approximate estimation in deep neural network models. *Cognition* **2021**, *215*, 104815. [CrossRef]

34. Yuan, X.; Ou, C.; Wang, Y.; Yang, C.; Gui, W. A Layer-Wise Data Augmentation Strategy for Deep Learning Networks and Its Soft Sensor Application in an Industrial Hydrocracking Process. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 3296–3305. [CrossRef] [PubMed]

35. Jurtz, V.I.; Johansen, A.R.; Nielsen, M.; Almagro Armenteros, J.J.; Nielsen, H.; Sønderby, C.K.; Winther, O.; Sønderby, S.K. An introduction to deep learning on biological sequence data: Examples and solutions. *Bioinformatics* **2017**, *33*, 3685–3690. [CrossRef] [PubMed]

36. Min, S.; Lee, B.; Yoon, S. Deep learning in bioinformatics. *Brief. Bioinform.* **2017**, *18*, 851–869. [CrossRef]

37. Laino, M.E.; Ammirabile, A.; Posa, A.; Cancian, P.; Shalaby, S.; Savevski, V.; Neri, E. The Applications of Artificial Intelligence in Chest Imaging of COVID-19 Patients: A Literature Review. *Diagnostics* **2021**, *11*, 1317. [CrossRef] [PubMed]

38. Wang, H.; Wang, L.; Lee, E.H.; Zheng, J.; Zhang, W.; Halabi, S.; Liu, C.; Deng, K.; Song, J.; Yeom, K.W. Decoding COVID-19 pneumonia: Comparison of deep learning and radiomics CT image signatures. *Eur. J. Nucl. Med. Mol. Imaging* **2021**, *48*, 1478–1486. [CrossRef]
39. Cao, C.; Liu, F.; Tan, H.; Song, D.; Shu, W.; Li, W.; Zhou, Y.; Bo, X.; Xie, Z. Deep Learning and Its Applications in Biomedicine. *Genom. Proteom. Bioinform.* **2018**, *16*, 17–32. [CrossRef]
40. Khan, S.H.; Hayat, M.; Porikli, F. Regularization of deep neural networks with spectral dropout. *Neural Netw.* **2019**, *110*, 82–90. [CrossRef] [PubMed]
41. Ma, J.; Yu, M.K.; Fong, S.; Ono, K.; Sage, E.; Demchak, B.; Sharan, R.; Ideker, T. Using deep learning to model the hierarchical structure and function of a cell. *Nat. Methods* **2018**, *15*, 290–298. [CrossRef]
42. Chen, C.; Shi, H.; Jiang, Z.; Salhi, A.; Chen, R.; Cui, X.; Yu, B. DNN-DTIs: Improved drug-target interactions prediction using XGBoost feature selection and deep neural network. *Comput. Biol. Med.* **2021**, *136*, 104676. [CrossRef]
43. Poernomo, A.; Kang, D.K. Biased Dropout and Crossmap Dropout: Learning towards effective Dropout regularization in convolutional neural network. *Neural Netw.* **2018**, *104*, 60–67. [CrossRef]
44. Li, H.; Weng, J.; Mao, Y.; Wang, Y.; Zhan, Y.; Cai, Q.; Gu, W. Adaptive Dropout Method Based on Biological Principles. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 4267–4276. [CrossRef] [PubMed]
45. Li, L.; Zhang, C.; Liu, S.; Guan, H.; Zhang, Y. Age Prediction by DNA Methylation in Neural Networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *3*, 1393–1402. [CrossRef] [PubMed]
46. Xie, J.; Ma, Z.; Lei, J.; Zhang, G.; Xue, J.H.; Tan, Z.H.; Guo, J. Advanced Dropout: A Model-free Methodology for Bayesian Dropout Optimization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, 1. [CrossRef] [PubMed]
47. Consistent Individualized Feature Attribution for Tree Ensembles. *arXive* **2019**, arXiv:1802.03888v3.
48. A Unified Approach to Interpreting Model Predictions. *arXive* **2017**, arXiv:1705.07874v2.
49. Rhodes, N.J.; O'Donnell, J.N.; Lizza, B.D.; McLaughlin, M.M.; Esterly, J.S.; Scheetz, M.H. Tree-Based Models for Predicting Mortality in Gram-Negative Bacteremia: Avoid Putting the CART before the Horse. *Antimicrob. Agents Chemother.* **2016**, *60*, 838–844. [CrossRef]
50. Parikh, S.A.; Gomez, R.; Thirugnanasambandam, M.; Chauhan, S.S.; De Oliveira, V.; Muluk, S.C.; Eskandari, M.K.; Finol, E.A. Decision Tree Based Classification of Abdominal Aortic Aneurysms Using Geometry Quantification Measures. *Ann. Biomed. Eng.* **2018**, *46*, 2135–2147. [CrossRef]
51. Freeman, C.; Kuli, D.; Basir, O. Feature-selected tree-based classification. *IEEE Trans. Cybern.* **2013**, *43*, 1990–2004. [CrossRef]
52. Vaid, A.; Somani, S.; Russak, A.J.; De Freitas, J.K.; Chaudhry, F.F.; Paranjpe, I.; Johnson, K.W.; Lee, S.J.; Miotto, R.; Richter, F.; et al. Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients With COVID-19 in New York City: Model Development and Validation. *J. Med. Internet Res.* **2020**, *22*, e24018. [CrossRef]
53. Li, X.; Ge, P.; Zhu, J.; Li, H.; Graham, J.; Singer, A.; Richman, P.S.; Duong, T.Q. Deep learning prediction of likelihood of ICU admission and mortality in COVID-19 patients using clinical variables. *PeerJ* **2020**, *8*, e10337. [CrossRef]
54. Wernly, B.; Heramvand, N.; Masyuk, M.; Rezar, R.; Bruno, R.R.; Kelm, M.; Niederseer, D.; Lichtenauer, M.; Hoppe, U.C.; Bakker, J.; et al. Acidosis predicts mortality independently from hyperlactatemia in patients with sepsis. *Eur. J. Intern. Med.* **2020**, *76*, 76–81. [CrossRef] [PubMed]
55. Kumar, V.; Karon, B.S. Comparison of measured and calculated bicarbonate values. *Clin. Chem.* **2008**, *54*, 1586–1587. [CrossRef] [PubMed]
56. Pourbagheri-Sigaroodi, A.; Bashash, D.; Fateh, F.; Abolghasemi, H. Laboratory findings in COVID-19 diagnosis and prognosis. *Clin. Chim. Acta* **2020**, *510*, 475–482. [CrossRef] [PubMed]
57. Evans, L.; Rhodes, A.; Alhazzani, W.; Antonelli, M.; Coopersmith, C.M.; French, C.; Machado, F.R.; McIntyre, L.; Ostermann, M.; Prescott, H.C.; et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021. *Crit. Care Med.* **2021**, *49*, e1063–e1143. [CrossRef] [PubMed]