



# Evaluating the Application of Machine Learning in Predicting the Mortality of Hospitalized COVID-19 Patients Using the Confusion Matrix and the Matthews Correlation Coefficient

Maryam Salari<sup>1</sup>, Seyed Masoud Sadati<sup>2</sup>, Alireza Sedaghat<sup>3</sup>, Bita Abbasi<sup>4</sup>, Seyed Amir Zamanpour<sup>5</sup>, Rozita Khodashahi<sup>6</sup>, Mostafa Davoudi<sup>1,\*</sup>

<sup>1</sup> Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran

<sup>2</sup> Center of Statistics and Information Technology Management, Imam Reza Hospital, Mashhad University of Medical Sciences, Mashhad, Iran

<sup>3</sup> Lung Disease Research Center, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

<sup>4</sup> Mashhad University of Medical Sciences, Mashhad, Iran

<sup>5</sup> Department of Medical Physics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

<sup>6</sup> Transplant Research Center, Clinical Research Institute, Mashhad University of Medical Sciences, Mashhad, Iran

\*Corresponding Author: Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran. Email: davoudim5@mums.ac.ir

Received: 26 August, 2024; Revised: 10 December, 2024; Accepted: 12 December, 2024

## Abstract

**Background:** The COVID-19 pandemic, which occurred between 2019 and 2023, posed a significant threat to global health. Its high transmissibility, the emergence of new variants, and the novel nature of the disease made treatment and control highly challenging.

**Objectives:** This study aimed to develop an algorithm for predicting the mortality of hospitalized COVID-19 patients using machine learning methods.

**Methods:** This cross-sectional study was conducted on 581 hospitalized COVID-19 patients. The approach integrated multi-model features derived from computed tomography (CT) scans and electronic health record (EHR) data. High-resolution computed tomography (HRCT) images were initially processed using the Pulmonary Toolkit package in MATLAB software. Subsequently, the extracted variables were entered into the model as predictive factors, alongside demographic characteristics, underlying conditions, and laboratory results of the patients. The machine learning model was developed using the AdaBoost method by incorporating demographic and laboratory data with HRCT features.

**Results:** In this study, 581 hospitalized COVID-19 patients were included. Among them, 199 (34.25%) patients died, while 382 (65.75%) recovered. According to the machine learning algorithm, the most effective variables for predicting COVID-19 mortality were lymphocyte variables, CRP, age, mean lung density, lung tissue percentage, RBC count, D-dimer levels, and emphysema. The MCC Index in this study was 0.73, and the area under the ROC curve was 0.96.

**Conclusions:** According to our results, the three variables with the greatest impact on predicting mortality in COVID-19 patients were related to HRCT findings, laboratory results, and patient age. Therefore, it is recommended that, given the high cost of HRCT, this diagnostic test should only be performed if other risk factors are identified in laboratory results. If necessary, HRCT should be conducted promptly.

**Keywords:** Machine Learning, COVID-19, Adaptive Boosting, HRCT, Laboratory Tests

## 1. Background

COVID-19 is an acute respiratory infectious disease caused by the virus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (1). In January 2020, the World Health Organization (WHO) declared the COVID-

19 outbreak a pandemic. The clinical outcomes of COVID-19 range from mild symptoms to severe complications and ultimately death, making it a significant global health concern (2, 3). The rapid spread of the disease has resulted in shortages of medical equipment and burnout among healthcare workers (4,

5). As of August 16, 2023, there have been 770,437,327 confirmed cases of COVID-19 globally, including 6,956,900 deaths (6). In the United States, from January 2020 until June 2023, there were 103,436,829 confirmed cases of COVID-19, with 1,127,152 deaths (6).

The main symptoms of COVID-19 include fever, cough, and shortness of breath, and the virus has high transmission and prevalence rates. Other symptoms may include tiredness, loss of taste or smell, muscle aches, chills, sore throat, runny nose, headache, chest pain, pink eye, nausea, vomiting, diarrhea, and rash (7). The severity of COVID-19 symptoms can range from very mild to severe. Some individuals may have no symptoms at all but can still spread the virus through asymptomatic transmission. The virus spreads via respiratory droplets released when someone coughs, sneezes, breathes, sings, or talks. COVID-19 is highly contagious and has led to a rapid pandemic that poses a serious threat to global public health (8, 9).

Due to the limited availability of diagnostic tests, accurately diagnosing COVID-19 remains one of the major challenges in managing this disease (10, 11). Alongside polymerase chain reaction (PCR) diagnostic tests, chest computed tomography (CT) scans are a significant diagnostic method for detecting the virus and monitoring the progression of the disease. Although chest CT scans may yield “false positives” in some cases, they remain a powerful tool for disease diagnosis. According to specialist reports, three types of abnormalities on CT scan images indicate COVID-19 infection: (1) Ground glass opacification, (2) consolidation, and (3) pleural effusion (12). Developing new tools for the improved detection of these irregularities in radiology images can greatly aid in controlling and managing COVID-19 (11).

Recently, the application of artificial intelligence and machine learning methods has been recognized as an efficient approach in the medical field. For example, Causey et al. reported an algorithm for predicting lung cancer using CT scan images and deep learning approaches, achieving an accuracy of 78% (13). Ardakani et al. developed eight machine learning models to distinguish COVID-19 from other non-COVID-19 lung diseases, achieving a ROC AUC of 0.994 for COVID-19 detection using their recent model (14).

## 2. Objectives

Given the importance of timely disease diagnosis, this study employed machine learning methods to predict the mortality of patients with COVID-19 infection.

## 3. Methods

### 3.1. Type of Study, Study Design, and Patient Selection

This cross-sectional study was conducted on 1,100 confirmed COVID-19 patients who were hospitalized at Imam Reza or Qaem hospitals under Mashhad University of Medical Sciences between December 2019 and December 2021. The 1,100 patients were selected using systematic random sampling. If a patient's high-resolution computed tomography (HRCT) information was unavailable, another patient was selected as a replacement. Demographic information, chronic disease history, imaging findings, vital signs, and laboratory results were collected from the patients' electronic medical records at the time of admission. The flowchart detailing patient selection and data collection is presented in Figure 1.

The patients' tissue and lung size values from CT scan images were analyzed using the Pulmonary Toolkit package in MATLAB software (Figure 2). Among the remaining 953 samples, 604 patients recovered, while 349 patients died. After incorporating laboratory information, 581 patients with complete data were included in the study.

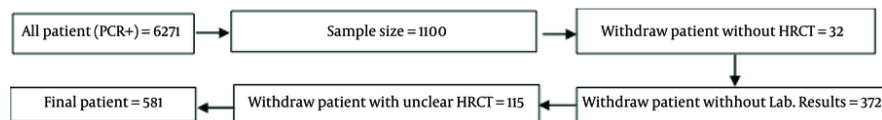
### 3.2. Inclusion and Exclusion Criteria

The inclusion criteria for this study were as follows: Patients who were hospitalized with a positive PCR test and had HRCT images available. Patients without HRCT images were excluded from the study.

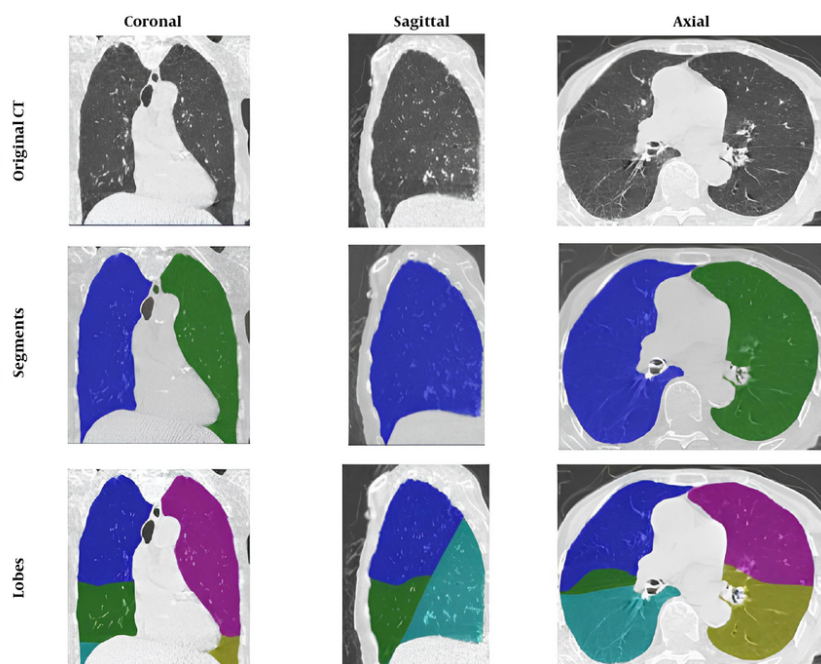
### 3.3. Statistics and Machine Learning Algorithm

Numerical variables were summarized using mean and standard deviation. To enhance the performance of data mining models and to determine the relationships between variables affecting COVID-19 mortality, *t*-tests, Mann-Whitney U tests, and chi-squared tests were employed. These tests were used to identify significant associations between variables and patient outcomes (death or recovery). A P-value < 0.05 was considered statistically significant. Variables with a significant relationship to the response variable were identified as risk factors.

In the machine learning model, the chi-squared ( $\chi^2$ ) feature selection algorithm was used to identify significant variables, accommodating both quantitative and qualitative variables. This algorithm is based on the  $\chi^2$  statistic. The  $\chi^2$  value for *r*, defined as the difference in *k* classes, is represented as follows (15).



**Figure 1.** Data selection scenario



**Figure 2.** An example of a lung and the corresponding derived lung. The first row shows the original CT scan. The second row illustrates the segmented binary mask using our segmentation pipeline. The third row shows the lobes.

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}}$$

$n_{ij}$ : Is the feature of  $j^{\text{th}}$  case

$$\mu_{ij} = \frac{n_{*j}n_{i*}}{n}$$

$n_{i*}$ : The number of  $i^{\text{th}}$  feature at all features

$n_{*j}$ : The number of samples in  $j^{\text{th}}$  class

$n$  = Sample size

In this study, Adaptive Boosting (AdaBoost) was used as a machine learning method to predict COVID-19-

related conditions, considering the type and quality of the data (Table 1). AdaBoost is based on decision tree algorithms and works by combining a high-accuracy predictor with variables that have relatively weaker accuracy (16, 17).

To evaluate the model's performance, 10-fold cross-validation was implemented. This statistical method for machine learning divides the dataset into training and validation sets across multiple iterations, ensuring that each data point is tested. Performance metrics, including accuracy, precision, recall, F-score, ROC AUC, and MCC, were calculated to assess the effectiveness of the predictive models (Table 2).

**Table 1.** Confusion Matrix

Variable	Predicted Values	
Actual values	Death (+)	Recover (-)
Death (+)	TP	FN
Recover (-)	FP	TN

**Table 2.** Performance Metrics Formulas

Formulas	Performance Metrics
$\frac{TP + TN}{TP + FP + TN + FN}$	Accuracy
$\frac{TP}{TP + FP}$	Precision
$\frac{TP}{TP + FN}$	Recall
$2 \times \frac{Precision \times sensitivity}{Precision + sensitivity}$	F-score
$MCC = \frac{T_N \times T_P - F_N \times F_P}{(T_P + F_P)(T_P + F_N)(T_N + F_P)(T_N + F_N)}$	Matthew's correlation coefficient

In summary, the eligibility criteria and statistical methods were as follows:

### 3.4. Eligibility Criteria

The eligibility criteria for inclusion in the study were:

- Patients must have been hospitalized with a confirmed diagnosis of COVID-19 via PCR.
- Availability of HRCT images was required; patients without these images were excluded.
- A total of 32 patients were excluded due to missing HRCT images. From the remaining 1,068 DICOM images, 115 samples were discarded due to unclear imaging. Ultimately, 581 patients with complete data were included in the analysis.

### 3.5. Statistical Methods

To analyze the data and determine relationships between variables affecting COVID-19 mortality, several statistical methods were employed:

- *t*-tests, Mann-Whitney tests, and chi-squared tests were used to identify significant relationships between variables and patient outcomes (death or recovery), with a significance level set at  $P < 0.05$ .
- Variables that demonstrated a significant relationship with mortality were classified as risk factors for prediction purposes.

- Adaptive Boosting (AdaBoost), based on decision tree algorithms, was utilized to enhance prediction accuracy by combining strong predictors with weaker ones.

- To evaluate the model's performance, 10-fold cross-validation was implemented. This method divides the dataset into training and validation sets across multiple iterations, ensuring that each data point is tested.

- Performance metrics, including accuracy, precision, recall, F-score, ROC AUC, and MCC, were calculated to assess the effectiveness of the predictive models.

### 3.6. Mitigation Strategies for Bias

Mitigation strategies for various types of biases in this study included:

- Using data from referral hospitals and employing systematic sampling methods to reduce selection bias.
- Involving expert clinicians and methodologists to minimize measurement biases.

The analysis may not fully account for confounding factors that could influence patient outcomes, such as variations in treatment protocols or differences in healthcare access among different populations. A total of 32 patients were excluded due to missing HRCT images, and an additional 115 samples were discarded due to unclear imaging. These exclusions could result in

**Table 3.** Demographic Characteristics, Comorbidities, Image Processing Result and Laboratory Finding on Admission <sup>a</sup>

Variables	Death (n = 199)	Non-death (n = 382)	P-Value
Age	66.57	57.34	< 0.001
<b>Gender</b>			0.31
Male	113 (56.7)	182 (47.6)	
Female	86 (43.2)	200 (52.3)	
<b>Comorbidities</b>			
Nausea	8 (4)	24 (6)	0.276
Cancer	8 (4)	2 (0.5)	0.002
Diabetes	27 (13)	60 (15)	0.493
Asthma	3 (1.5)	4 (1)	0.629
Heart disease	10 (5)	30 (7)	0.201
Chronic kidney disease	4 (2)	5 (1)	0.516
Chronic lung disease	5 (2)	5 (1)	0.29
Hypertension	38 (19)	79 (20)	0.651
PO <sub>2</sub>	83.7	87.7	< 0.001
<b>Image processing result</b>			
Percent of air	61.18	65.22	< 0.001
Volume of air, cm <sup>3</sup>	1490.27	1698.42	0.005
Percent of emphysema	10.97	11.58	0.403
Mean density, HU	-611.81	-652.22	< 0.001
Percent of tissue	38.81	34.77	< 0.001
Emphysema	896.49	911.89	0.027
<b>Laboratory finding on admission</b>			
White blood cell, × 1000/mL	8.41	7.24	0.236
Red blood cell, × 1000/mL	9.81	6.26	0.003
LDL	104.2	68.75	0.338
Ferritin, ng/mL	689.08	541.74	0.027
FBS	156.10	209.57	0.069
D-dimer, ng/mL	325.8	167.2	< 0.001
C-reactive protein, mg/dL	10.8	6.6	< 0.001
Percent of Lymphocytes	14.15	9.65	< 0.001

<sup>a</sup> Values are expressed as No. (%) unless otherwise indicated.

the loss of potentially relevant data and may impact the generalizability of the findings.

While the study provides valuable insights into COVID-19 patient outcomes, caution should be exercised when applying its findings to broader populations due to differences in demographics, healthcare practices, and the evolving treatment landscape.

#### 4. Results

The study was conducted on 581 patients, of whom 295 were male with an average age of 50.3 years, and 286 were female with an average age of 50 years. Of these, 199 patients were in the mortality group, and 382 were in the recovery group. The descriptive statistics are presented in [Table 3](#).

The average age of deceased patients was 66.57 years, compared to 57.34 years for recovered patients. In the comorbidities subgroup analysis, only patients with cancer showed a significant difference (P-value = 0.002, [Table 3](#)). Additionally, the comparison of the mean SpO<sub>2</sub> Index between deceased and recovered patients was statistically significant (P-value = 0.002).

From the image processing results, the mean lung density, the percentage of air in the lungs, and the volume of air in the lungs were statistically different between the deceased and recovered groups (P-value < 0.001, P-value < 0.001, and P-value = 0.005, respectively). Furthermore, the emphysema Index in the recovery group was significantly higher than in the deceased group (P-value = 0.002).

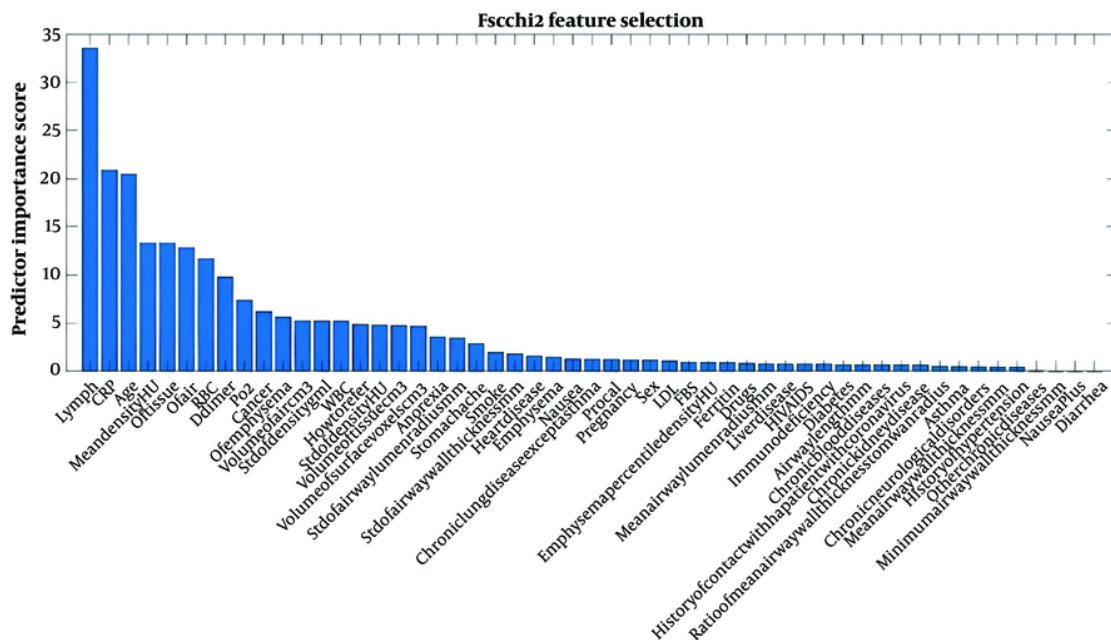


Figure 3. Feature selection using the fscchi2 method

In laboratory findings, there were significant differences in red blood cell (RBC) counts, lymphocyte levels, C-reactive protein (CRP), ferritin, and D-dimer levels between deceased and recovered patients.

The Adaptive Boosting (AdaBoost) model was fitted to the data to predict treatment outcomes, and the three variables with the most significant impact on prediction are shown in Figure 3. In the final model, 10 variables—including lymphocytes, CRP, age group, mean tissue density, RBC, D-dimer,  $pO_2$ , cancer, and emphysema—were identified as having the most significant impact on prediction and were included in the analysis.

After fitting the model to predict treatment outcomes, the ROC curve was plotted to evaluate the model, yielding an AUC of 0.96 (Figure 4). Table 4 presents the confusion matrix for the AdaBoost model in predicting the outcomes of hospital care for COVID-19 inpatients. The evaluation metrics of the model are displayed in Table 5. The accuracy and precision of the model were 0.88 and 0.89, respectively. Predictive models like this one aim to maximize the agreement between predicted and actual values regarding recovery and mortality. Matthew's correlation coefficient (MCC) showed a value of 0.73 (Table 5).

## 5. Discussion

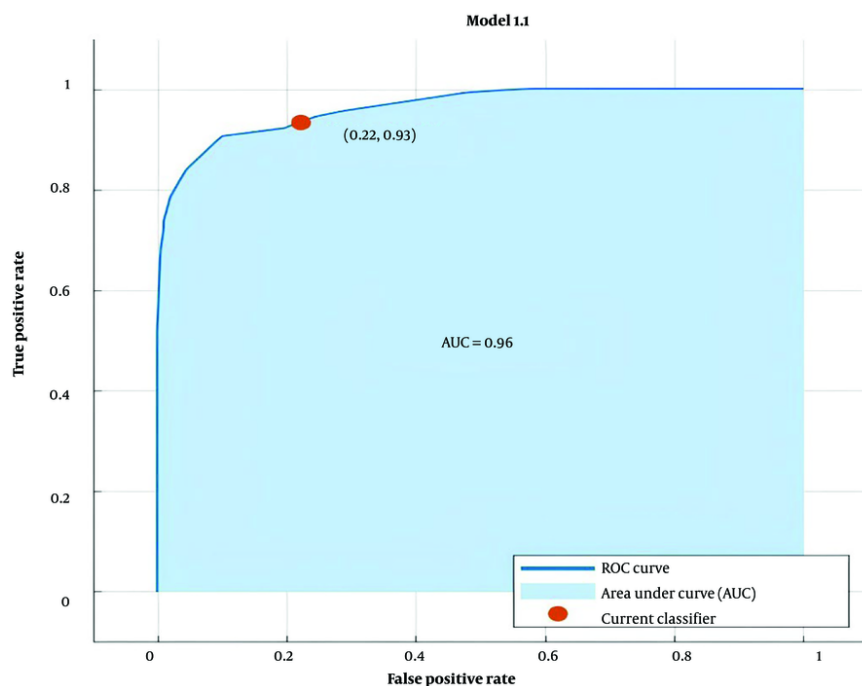
This study presents a retrospective analysis of patient data to predict the mortality of COVID-19 patients hospitalized in referral hospitals between 2010 and 2021. Machine learning algorithms were applied to predict disease outcomes based on clinical data from hospitalized patients.

Lai et al. used the Adaptive Boosting algorithm to identify the most effective variables for predicting mortality in COVID-19 patients. Their findings revealed that lymphocyte counts were significantly lower in patients with severe COVID-19 compared to those with mild cases (18).

Lymphocyte count and CRP are two important variables in predicting the risk of death in patients with COVID-19. Several studies have shown that lymphocyte count serves as a universal predictor of health outcomes in COVID-19 patients (19).

Winderadi et al. have indicated that CRP, as an acute-phase protein, is an effective marker for predicting severe COVID-19 (20). In a meta-analysis study, it was demonstrated that CRP is a significant variable in distinguishing between severe and mild cases of COVID-19 (21).

In the present study, we found that RBC was an effective variable for predicting the risk of death in



**Figure 4.** Receiver operating characteristic (ROC) curve for evaluating the predictive power of treatment outcomes in patients with COVID-19

**Table 4.** Confusion Matrix for AdaBoost Model for Predicting the Outcome of Hospital Care of COVID-19 Inpatient

Variables	Predicted Survival	Predicted Death
Actual survival	357	25
Actual death	44	155

**Table 5.** Indices for the AdaBoost Model for Predicting the Outcome of Hospital Care for COVID-19 Inpatients

Index	AdaBoost
Accuracy	0.88
Precision	0.89
F-measure	0.84
Recall	93.1
Matthew's correlation coefficient	0.73

COVID-19 patients. Hemoglobin in RBCs is considered an important biomarker, reflecting oxygen levels in the blood and serving as a significant variable in predicting COVID-19 mortality (22). Thomas et al. showed that RBC counts were significantly higher in COVID-19 patients compared to healthy individuals (23).

Additionally, age has been identified as a crucial variable for predicting COVID-19 mortality (24, 25). Bonanad et al. conducted a meta-analysis of 611,583 COVID-19 patients across five continents to investigate mortality rates among different age groups. They found that the mortality rate for individuals under 50 years old was 1.1%, and this rate increased with age, peaking in

individuals aged 80 years or older (26). Another study found that individuals aged 55-64 years had an 8.1-fold higher COVID-19 mortality rate than those under 55 years of age (27). These findings suggest that age is a significant predictor of COVID-19 mortality. As age increases, the mortality rate also rises, with the highest mortality rates observed in patients aged 80 years and above (24).

Lyu *et al.* aimed to evaluate the severity of COVID-19 based on HRCT images. They found that the mean lung density, measured on the HU Scale, was higher in patients with severe COVID-19 compared to healthy individuals (28). In our study, the mean lung density in deceased individuals was also found to be higher than in those who recovered. Notably, the diagnostic value of CT scanning in assessing lung density has already been well-established and is considered preferable to other subjective visual examinations (29).

The data suggest that lung density is a potential imaging tool for assessing the severity of COVID-19, and its results can be valuable for identifying patients at risk of severe disease progression (30). However, further studies are necessary to validate the clinical utility of lung density analysis in managing COVID-19.

Additionally, we observed that the average D-dimer level was significantly lower in recovered individuals compared to deceased patients ( $P$ -value = 0.001). D-dimer is a blood biomarker that plays a critical role in predicting outcomes for patients with COVID-19 (31). One study indicated that the mean D-dimer level in patients with mild COVID-19 was approximately one-sixth of that in patients with severe disease (32).

It has also been demonstrated that patients with malignancies are at a higher risk of COVID-19 infection and severe complications due to their immunocompromised state (33). Similarly, other studies have reported an increased rate of COVID-19-associated mortality among cancer patients (34, 35).

The risk of severe COVID-19 outcomes increases with age, and patients with malignant tumors are at a higher risk for severe illness due to their underlying medical conditions (36). During the COVID-19 pandemic, cancer patients have had limited access to medical facilities and services, which has increased the likelihood and severity of their conditions (37). In our study, a significant difference was observed in the proportion of cancer patients between the deceased and recovered groups ( $P$ -value = 0.02). Patients with malignancies are at higher risk for severe complications and mortality from COVID-19 due to their immunocompromised state and underlying medical conditions. Vaccination has been shown to help reduce deaths and severe illness

from COVID-19, as well as to decrease transmission in these patients (38).

In recent studies, predicting the severity and mortality of COVID-19 has been a major focus. Several studies have explored the relationship between COVID-19 and mortality, including excess mortality due to COVID-19, as well as machine learning models to predict mortality and critical events in COVID-19 patients. In a study by Akhtar *et al.*, 10 machine learning algorithms were used to predict COVID-19 infection based on CBC results (39). According to their results, the highest accuracy (100%) in predicting infection was achieved by three algorithms: Random Forest, K Nearest Neighbor (KNN), and kStar. These findings suggest that machine learning algorithms can be useful in predicting COVID-19 infection based on CBC results. Further research is needed to establish the clinical utility of these algorithms in managing COVID-19. Moulai *et al.* conducted a study on 1500 COVID-19 patients to predict mortality using various machine learning models. Their results showed that the ML and RF methods had the highest accuracy (> 80%) (1). In another study, Zakariaee *et al.* assessed the performance of four machine learning algorithms (LR, RF, SVM, and XGBoost) and found that XGBoost had the best performance in terms of AUC (40).

Schiaffino *et al.* conducted a study on 897 hospitalized COVID-19 patients to predict in-hospital mortality using HRCT scans. The algorithms used in this study were Support Vector Machine (SVM) and multi-layer perceptron (MLP). The area under the ROC curve for the SVM and MLP models was 0.74 and 0.84, respectively (41). Nuthalapati *et al.* used deep learning methods to predict mortality or hospitalization in the intensive care unit (ICU) for COVID-19 patients. Other variables, such as HRCT images and electronic health record (HER) data, were used in this study. They found that the normal lung volume, normal lung percentage (NLperc), muscle volume, fat volume, muscle-fat ratio, age, sex, and lesion percentage were the most important variables for predicting mortality and ICU hospitalization. The area under the ROC curve was approximately 0.77 (42). Other studies have also explored the use of deep learning algorithms in analyzing body composition on CT scans to predict outcomes in COVID-19 patients. In this context, Zhang *et al.* (as cited by Nachit *et al.*) used a deep learning algorithm to analyze body composition on CT scans and found that myosteosis was a key predictor of mortality in asymptomatic adults (43). These findings suggest that deep learning algorithms can be useful in predicting outcomes in COVID-19 patients based on body



composition analysis. Further research is needed to establish the clinical utility of these algorithms in COVID-19 management.

Machine learning algorithms have been used in many studies to predict COVID-19 mortality. Some studies have used only clinical features, while others have incorporated radiological features as well. The selection of ML algorithms was based on related studies in the field and the quality of the selected dataset. The most commonly used algorithms were SVM, MLP, RF, KNN, and kStar. The performance of the models was evaluated using metrics derived from the confusion matrix, such as AUC and MCC. Important predictors for COVID-19 patient mortality included lymphocyte count, CRP, age, mean lung density, lung tissue percentage, RBC, D-Dimer, and emphysema. The AUC of the models ranged from 0.74 to 0.96. Some studies also used deep learning techniques and EHR data to predict mortality or hospitalization in COVID-19 patients.

In most studies, only the ROC curve, which is a function of the accuracy of predictions, is reported, typically yielding good results. However, in the present study, the agreement of the 4 cells in the contingency table was calculated using MCC. This showed that, although the model may perform well in predicting patient improvement, it may not perform as well in predicting patient mortality, which is the primary concern. For example, in the Gong study, it was shown that all confusion matrix indices focus solely on false positives, while only the MCC Index takes into account both false positives and false negatives (44).

### 5.1. Conclusions

The main limitations of our study include the possibility that our analysis may not fully account for confounding factors that could influence patient outcomes, such as variations in treatment protocols or differences in healthcare access across different populations. We suggest that simulation studies should be used to enhance understanding and create appropriate indices for machine learning methods, which can be selected based on the type of data. The three variables with the greatest impact on predicting mortality in COVID-19 patients were related to laboratory results, with age being the next most significant variable. Therefore, we recommend that, due to cost, HRCT should only be performed if risk factors are observed in laboratory results, and if necessary, HRCT should be performed promptly.

### Footnotes

**Authors' Contribution:** Study concept and design: M. D. and M. S.; Acquisition of data: S. M. S.; Analysis and interpretation of data: M. S.; Drafting of the manuscript: R. Kh. and B. A.; Critical revision of the manuscript for important intellectual content: A. S. and R. Kh.; Statistical analysis: M. D. and S. A. Z.; Administrative, technical, and material support: M. D. and M. S.; Study supervision: M. S.

**Conflict of Interests Statement:** The authors declare that they have no competing interest.

**Data Availability:** The data presented in this study are uploaded as a supplementary file during submission and are available to readers upon request.

**Ethical Approval:** In this study, patient ID numbers were used, and no personal patient information was collected. This study was approved by the Ethics Committee of Mashhad University of Medical Sciences under the ethics code [IR.MUMS.REC.1401.042](https://doi.org/10.1186/s12911-021-01742-0).

**Funding/Support:** This study was supported by the Mashhad University of Medical Sciences.

### References

- Moulaei K, Shanbehzadeh M, Mohammadi-Taghiabad Z, Kazemi-Arpanahi H. Comparing machine learning algorithms for predicting COVID-19 mortality. *BMC Med Inform Decis Mak*. 2022;**22**(1):2. [PubMed ID: [34983496](https://pubmed.ncbi.nlm.nih.gov/34983496/)]. [PubMed Central ID: [PMC8724649](https://pubmed.ncbi.nlm.nih.gov/PMC8724649/)]. <https://doi.org/10.1186/s12911-021-01742-0>.
- Gupta N, Dhamija S, Patil J, Chaudhari B. Impact of COVID-19 pandemic on healthcare workers. *Ind Psychiatry J*. 2021;**30**(Suppl 1):S282-4. [PubMed ID: [34908710](https://pubmed.ncbi.nlm.nih.gov/34908710/)]. [PubMed Central ID: [PMC8611576](https://pubmed.ncbi.nlm.nih.gov/PMC8611576/)]. <https://doi.org/10.4103/0972-6748.328830>.
- OECD. *Health at a Glance 2021: OECD Indicators*. Paris: OECD Publishing; 2021.
- Leo CG, Sabina S, Tumolo MR, Bodini A, Ponzini G, Sabato E, et al. Burnout Among Healthcare Workers in the COVID 19 Era: A Review of the Existing Literature. *Front Public Health*. 2021;**9**:750529. [PubMed ID: [34778184](https://pubmed.ncbi.nlm.nih.gov/34778184/)]. [PubMed Central ID: [PMC8585922](https://pubmed.ncbi.nlm.nih.gov/PMC8585922/)]. <https://doi.org/10.3389/fpubh.2021.750529>.
- Health UDO; H. Services. *Impact of the COVID-19 pandemic on the hospital and outpatient clinician workforce*. 2022. Available from: <https://aspe.hhs.gov/reports/covid-19-health-care-workforce>.
- WHO. *Number of COVID-19 cases reported to WHO*. Geneva, Switzerland: WHO; 2024. Available from: <https://data.who.int/dashboards/covid19/cases?n=c>.
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;**395**(10223):497-506. [PubMed ID: [31986264](https://pubmed.ncbi.nlm.nih.gov/31986264/)]. [PubMed Central ID: [PMC7159299](https://pubmed.ncbi.nlm.nih.gov/PMC7159299/)]. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5).
- Cascella M, Rajnik M, Aleem A, Dulebohn SC, Di Napoli R. Features, Evaluation, and Treatment of Coronavirus (COVID-19). *StatPearls*. Treasure Island (FL); 2024.
- Sadeghi Dousari A, Taati Moghadam M, Satarzadeh N. COVID-19 (Coronavirus Disease 2019): A New Coronavirus Disease. *Infect Drug*

- Resist. 2020;**13**:2819-28. [PubMed ID: 32848431]. [PubMed Central ID: PMC7429403]. <https://doi.org/10.2147/IDR.S259279>.
10. Chu DKW, Pan Y, Cheng SMS, Hui KPY, Krishnan P, Liu Y, et al. Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia. *Clin Chem*. 2020;**66**(4):549-55. [PubMed ID: 32031583]. [PubMed Central ID: PMC7108203]. <https://doi.org/10.1093/clinchem/hvaa029>.
  11. Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect Dis*. 2020;**20**(4):425-34. [PubMed ID: 32105637]. [PubMed Central ID: PMC7159053]. [https://doi.org/10.1016/S1473-3099\(20\)30086-4](https://doi.org/10.1016/S1473-3099(20)30086-4).
  12. Song F, Shi N, Shan F, Zhang Z, Shen J, Lu H, et al. Emerging 2019 Novel Coronavirus (2019-nCoV) Pneumonia. *Radiology*. 2020;**297**(3): E346. [PubMed ID: 33196374]. [PubMed Central ID: PMC8906333]. <https://doi.org/10.1148/radiol.2020209021>.
  13. Causey JL, Guan Y, Dong W, Walker K, Qualls JA, Prior F, et al. Lung cancer screening with low-dose CT scans using a deep learning approach. *arXiv*. 2019; **Preprint**. <https://doi.org/10.48550/arXiv.1906.00240>.
  14. Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. *Comput Biol Med*. 2020;**121**:103795. [PubMed ID: 32568676]. [PubMed Central ID: PMC7190523]. <https://doi.org/10.1016/j.compbiomed.2020.103795>.
  15. Liu H, Setiono R. A probabilistic approach to feature selection - a filter solution. *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*. Bari, Italy. Morgan Kaufmann Publishers Inc; 1996. 319-327 p.
  16. Kearns M, Valiant L. Cryptographic limitations on learning Boolean formulae and finite automata. *J ACM*. 1994;**41**(1):67-95. <https://doi.org/10.1145/174644.174647>.
  17. Kivinen J, Warmuth MK. Boosting as entropy projection. *Proceedings of the twelfth annual conference on Computational learning theory*. 1999. p. 134-44.
  18. Lai KL, Hu FC, Wen FY, Chen JJ. Lymphocyte count is a universal predictor of health outcomes in COVID-19 patients before mass vaccination: A meta-analytical study. *J Glob Health*. 2022;**12**:5041. [PubMed ID: 36112520]. [PubMed Central ID: PMC9480861]. <https://doi.org/10.7189/jogh.12.05041>.
  19. Wang S, Sheng Y, Tu J, Zhang L. Association between peripheral lymphocyte count and the mortality risk of COVID-19 inpatients. *BMC Pulm Med*. 2021;**21**(1):55. [PubMed ID: 33573626]. [PubMed Central ID: PMC7877317]. <https://doi.org/10.1186/s12890-021-01422-9>.
  20. Windradi C, Asmarawati TP, Rosyid AN, Marfiani E, Mahdi BA, Martani OS, et al. Hemodynamic, Oxygenation and Lymphocyte Parameters Predict COVID-19 Mortality. *Pathophysiology*. 2023;**30**(3):314-26. [PubMed ID: 37606387]. [PubMed Central ID: PMC10443272]. <https://doi.org/10.3390/pathophysiology30030025>.
  21. Kermali M, Khalsa RK, Pillai K, Ismail Z, Harky A. The role of biomarkers in diagnosis of COVID-19 - A systematic review. *Life Sci*. 2020;**254**:117788. [PubMed ID: 32475810]. [PubMed Central ID: PMC7219356]. <https://doi.org/10.1016/j.lfs.2020.117788>.
  22. Russo A, Tellone E, Barreca D, Ficarra S, Lagana G. Implication of COVID-19 on Erythrocytes Functionality: Red Blood Cell Biochemical Implications and Morpho-Functional Aspects. *Int J Mol Sci*. 2022;**23**(4). [PubMed ID: 35216286]. [PubMed Central ID: PMC8878454]. <https://doi.org/10.3390/ijms23042171>.
  23. Thomas T, Stefanoni D, Dzieciatkowska M, Issaian A, Nemkov T, Hill RC, et al. Evidence of Structural Protein Damage and Membrane Lipid Remodeling in Red Blood Cells from COVID-19 Patients. *J Proteome Res*. 2020;**19**(11):4455-69. [PubMed ID: 33103907]. [PubMed Central ID: PMC7640979]. <https://doi.org/10.1021/acs.jproteome.0c00606>.
  24. Henkens M, Raafs AG, Verdonschot JA, Linschoten M, van Smeden M, Wang P, et al. Age is the main determinant of COVID-19 related in-hospital mortality with minimal impact of pre-existing comorbidities, a retrospective cohort study. *BMC Geriatr*. 2022;**22**(1):184. [PubMed ID: 35247983]. [PubMed Central ID: PMC8897728]. <https://doi.org/10.1186/s12877-021-02673-1>.
  25. Chatterjee A, Wu G, Primakov S, Oberije C, Woodruff H, Kubben P, et al. Can predicting COVID-19 mortality in a European cohort using only demographic and comorbidity data surpass age-based prediction: An externally validated study. *PLoS One*. 2021;**16**(4): e0249920. [PubMed ID: 33857224]. [PubMed Central ID: PMC8049248]. <https://doi.org/10.1371/journal.pone.0249920>.
  26. Bonanad C, Garcia-Blas S, Tarazona-Santabalbina F, Sanchis J, Bertomeu-Gonzalez V, Facila L, et al. The Effect of Age on Mortality in Patients With COVID-19: A Meta-Analysis With 611,583 Subjects. *J Am Med Dir Assoc*. 2020;**21**(7):915-8. [PubMed ID: 32674819]. [PubMed Central ID: PMC7247470]. <https://doi.org/10.1016/j.jamda.2020.05.045>.
  27. Adjei S, Hong K, Molinari NM, Bull-Otterson L, Ajani UA, Gundlapalli AV, et al. Mortality Risk Among Patients Hospitalized Primarily for COVID-19 During the Omicron and Delta Variant Pandemic Periods - United States, April 2020-June 2022. *MMWR Morb Mortal Wkly Rep*. 2022;**71**(37):1182-9. [PubMed ID: 36107788]. [PubMed Central ID: PMC9484808]. <https://doi.org/10.15585/mmwr.mm7137a4>.
  28. Lyu P, Liu X, Zhang R, Shi L, Gao J. The Performance of Chest CT in Evaluating the Clinical Severity of COVID-19 Pneumonia: Identifying Critical Cases Based on CT Characteristics. *Invest Radiol*. 2020;**55**(7):412-21. [PubMed ID: 32304402]. [PubMed Central ID: PMC7173027]. <https://doi.org/10.1097/RLI.0000000000000689>.
  29. Yousef HA, Moussa EMM, Abdel-Razek MZM, El-Kholy MMSA, Hasan LHS, El-Sayed AEA, et al. Automated quantification of COVID-19 pneumonia severity in chest CT using histogram-based multi-level thresholding segmentation. *Egypt J Radiol Nucl Med*. 2021;**52**(1). <https://doi.org/10.1186/s43055-021-00602-1>.
  30. Bressemer KK, Adams LC, Albrecht J, Petersen A, Thiess HM, Niehues A, et al. Is lung density associated with severity of COVID-19? *Pol J Radiol*. 2020;**85**:e600-6. [PubMed ID: 33204375]. [PubMed Central ID: PMC7654311]. <https://doi.org/10.5114/pjr.2020.100788>.
  31. Ye W, Chen G, Li X, Lan X, Ji C, Hou M, et al. Dynamic changes of D-dimer and neutrophil-lymphocyte count ratio as prognostic biomarkers in COVID-19. *Respir Res*. 2020;**21**(1):169. [PubMed ID: 32620118]. [PubMed Central ID: PMC7332531]. <https://doi.org/10.1186/s12931-020-01428-7>.
  32. Yu HH, Qin C, Chen M, Wang W, Tian DS. D-dimer level is associated with the severity of COVID-19. *Thromb Res*. 2020;**195**:219-25. [PubMed ID: 32777639]. [PubMed Central ID: PMC7384402]. <https://doi.org/10.1016/j.thromres.2020.07.047>.
  33. Cona MS, Riva A, Dalu D, Gabrieli A, Fasola C, Lipari G, et al. Clinical efficacy of the first two doses of anti-SARS-CoV-2 mRNA vaccines in solid cancer patients. *Cancer Med*. 2023;**12**(12):12967-74. [PubMed ID: 37114577]. [PubMed Central ID: PMC10315797]. <https://doi.org/10.1002/cam4.5968>.
  34. Gupta K, Gandhi S, Mebane A3, Singh A, Vishnuvardhan N, Patel E. Cancer patients and COVID-19: Mortality, serious complications, biomarkers, and ways forward. *Cancer Treat Res Commun*. 2021;**26**:100285. [PubMed ID: 33360669]. [PubMed Central ID: PMC7832265]. <https://doi.org/10.1016/j.ctarc.2020.100285>.
  35. Wang H, Zhang L. Risk of COVID-19 for patients with cancer. *Lancet Oncol*. 2020;**21**(4): e181. [PubMed ID: 32142621]. [PubMed Central ID: PMC7129735]. [https://doi.org/10.1016/S1470-2045\(20\)30149-2](https://doi.org/10.1016/S1470-2045(20)30149-2).
  36. Centers for Disease Control and Prevention. *Underlying medical conditions associated with higher risk for severe COVID-19 : information for healthcare providers*. Georgia, US: Centers for Disease Control and

- Prevention; 2022. Available from: <https://stacks.cdc.gov/view/cdc/118293>.
37. Liu C, Zhao Y, Okwan-Duodu D, Basho R, Cui X. COVID-19 in cancer patients: risk, clinical features, and management. *Cancer Biol Med*. 2020;**17**(3):519-27. [PubMed ID: 32944387]. [PubMed Central ID: PMC7476081]. <https://doi.org/10.20892/j.issn.2095-3941.2020.0289>.
  38. WHO. *Statement for healthcare professionals: How COVID-19 vaccines are regulated for safety and effectiveness (Revised March 2022)*. Geneva, Switzerland: WHO; 2022. Available from: <https://www.who.int/news/item/17-05-2022-statement-for-healthcare-professionals-how-covid-19-vaccines-are-regulated-for-safety-and-effectiveness>.
  39. Akhtar A, Akhtar S, Bakhtawar B, Kashif AA, Aziz N, Javeid MS. COVID-19 Detection from CBC using Machine Learning Techniques. *Int J Innov Technol Manag*. 2021;**1**(2):65-78. <https://doi.org/10.54489/ijtim.v1i2.22>.
  40. Zakariaee SS, Abdi AI, Naderi N, Babashahi M. Prognostic significance of chest CT severity score in mortality prediction of COVID-19 patients, a machine learning study. *Egypt J Radiol Nucl Med*. 2023;**54**(1). <https://doi.org/10.1186/s43055-023-01022-z>.
  41. Schiaffino S, Codari M, Cozzi A, Albano D, Ali M, Arioli R, et al. Machine Learning to Predict In-Hospital Mortality in COVID-19 Patients Using Computed Tomography-Derived Pulmonary and Vascular Features. *J Pers Med*. 2021;**11**(6). [PubMed ID: 34204911]. [PubMed Central ID: PMC8230339]. <https://doi.org/10.3390/jpm11060501>.
  42. Nuthalapati SV, Vizcaychipi M, Shah P, Chudzik P, Leow CH, Yousefi P, et al. Using Deep Learning-based Features Extracted from CT scans to Predict Outcomes in COVID-19 Patients. *arXiv*. 2022;**Preprint**. <https://doi.org/10.48550/arXiv.2205.05009>.
  43. Nachit M, Horsmans Y, Summers RM, Leclercq IA, Pickhardt PJ. AI-based CT Body Composition Identifies Myosteatorsis as Key Mortality Predictor in Asymptomatic Adults. *Radiology*. 2023;**307**(5). e222008. [PubMed ID: 37191484]. [PubMed Central ID: PMC10315523]. <https://doi.org/10.1148/radiol.222008>.
  44. Gong M. A Novel Performance Measure for Machine Learning Classification. *Int. J. Manag. Inf. Technol*. 2021;**13**(1):11-9. <https://doi.org/10.5121/ijmit.2021.13101>.