Kurdish Studies
April 2024
Volume: 12, No: 2, pp 7025-7033
ISSN: 2051-4883 (Print) | ISSN 2051-4891 (Online)
www.KurdishStudies.net

DOI: 10.53555/ks.v12i2.3966

Clinical Profiling And Disease Severity Prediction Of COVID-19 By Multinomial Regression Analysis

Mahnoor Khan[1,2*], Awais Altaf[1], Syed Zeeshan Haider Naqvi[1], Tahir Maqbool[1], Hafiz Muhammad Hammad[3], Muhammad Abdullah Tanveer[3]

[1]Institute of Molecular Biology and Biotechnology, The University of Lahore, Lahore, Pakistan;
[2]Federal Postgraduate Medical Institute, Shaikh Zayed Hospital, Lahore, Pakistan;
[3]School of Biochemistry and Biotechnology, University of the Punjab, Lahore, Pakistan;

***Corresponding Author:** Mahnoor Khan
*Email: drmahnoorshoaib@gmail.com

Abstract: *Background and Objectives*: The biochemical evaluation of patients infected with SARS-CoV-2 is crucial for their on-site stratification, accurate assessment, and subsequent management. This cross-sectional study aims to evaluate the biochemical markers in patients of both genders with control (normal), ward (moderate), and ICU (severe) COVID-19. Furthermore, using the same dataset, disease severity prediction by multinomial regression analysis was performed to identify the major predictors of disease progression in COVID-19. *Materials and Methods:* Biochemical evaluations of all study subjects were performed using an automated chemistry analyzer. The statistical analysis was performed using the R programming language, while the multinomial regression analysis was performed using Python. Confusion matrix and ROC analysis were used to evaluate the predictive performance of this model. Cross-validation of the model was done to ensure that it generalizes well with new and unseen data. *Results:* Among ward patients, significantly increased levels of LDH ($p = 0.0028$) and CRP ($p = 0.0138$) were observed in females compared to males. In contrast, among ICU patients, CRP levels ($p = 0.0309$) rose more in males than in females. Moreover, according to the multinomial regression analysis, AST, IL-6, Direct Bilirubin, LDH, Neutrophils, ALT, and CRP were observed to be the major predictors of disease severity. The mean AUC obtained after 5-fold cross-validation was 0.90 ± 0.06. *Conclusion:* The biochemical analysis revealed CRP as the most important predictor of disease severity in males, whereas in females, LDH emerged as the most important marker of disease severity. Accordingly, AST, IL-6, Direct bilirubin, Neutrophils, LDH, ALT, and CRP were observed to have the most impact on the predictability of the disease status, as evident by the multinomial regression analysis. The high mean AUC value obtained after cross-validation of the model indicates its excellent ability to determine the disease severity status in COVID-19 patients.

Keywords: COVID-19; clinical profiling; biochemical evaluation; multinomial logistic regression; gender based evaluation in COVID-19

Introduction

One of the most infectious viral diseases of this century emerged in China in December 2019. The Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was identified as the causative agent responsible for the emergence of Coronavirus disease, later labelled as COVID-19. Within a short period, it became a major health concern due to the extremely high number of people affected worldwide. As of now, nearly 7 million deaths have been reported globally, with a much greater figure of morbidity related to this pandemic [1]. The disease has a diverse clinical outlook, ranging from the patient being totally asymptomatic to mild, moderate, severe, and critical forms of the disease. The most commonly occurring symptoms include cough, fever, diarrhea, dyspnea, and involvement of multiple organs. Mostly, the patients with a mild form of the disease usually recover, but some develop serious complications that include respiratory failure, which is the most prevalent one, followed by renal, cardiac, and neurological dysfunction [2]. Detection of this viral infection in its earliest phase, followed by immediate isolation and effective management, remains the most crucial aspect of management to reduce the rates of transmission and severity of the disease [3].

Several studies have highlighted the importance of clinical profiling in predicting the prognosis of this viral disease. The biochemical markers, including Lactate dehydrogenase (LDH), C-reactive protein (CRP), Ferritin, Interleukin-6 (IL-6), D-dimer, and Troponin I are considered vital in this regard [4,5,6,7]. Along with these inflammatory markers, other organ function markers, including Alanine aminotransferase (ALT) and Aspartate aminotransferase (AST), Blood urea nitrogen (BUN), and Creatinine, have also been reported as major predictors of disease progression in COVID-19 [8]. CRP is an established inflammatory marker, whereas LDH is a well-known marker for both inflammation and tissue damage [9]. A comprehensive analysis of all these circulating biochemical biomarkers may reveal effective screening tools for accurate diagnosis and prognosis.

Interestingly, various studies have undertaken biochemical investigations in people affected by SARS-CoV-2, but gender-based information from this perspective is scarcely available. In this study, therefore, we measured, analyzed, and reported biochemical markers in both genders variably affected by SARS-CoV-2. This was done to reveal the significant differences

between the biomarkers of the two genders, which may prove fruitful in devising new and more effective gender-based diagnostic and management strategies.
Another objective of this study was to build and validate a multinomial regression predictive model for clinical practice, based on the given data. This involves identifying different biochemical variables that can be used to categorize patients as normal, moderately affected, or severely affected by COVID-19. The aim is to improve disease severity prediction in COVID-19 and demonstrate the reliability of this model to facilitate informed decision-making in patient management [10].

Materials and Methods
This cross-sectional study was conducted at the Biochemistry Department of the Institute of Molecular Biology and Biotechnology (IMBB) at The University of Lahore, Pakistan. It was approved by the IMBB Ethical Review Committee, University of Lahore (IMBB/BBBC/22/120). Samples were collected from the local hospitals of Lahore. All the data acquired was kept confidential.
A total of 60 participants aged 18 years or older were recruited for this study after obtaining their written and informed consent. Out of these 60 subjects, 20 were diagnosed with COVID-19 patients admitted to the ward, 20 were diagnosed with COVID-19 patients admitted to the intensive care unit (ICU), and 20 were healthy individuals taken as controls. All groups had equal numbers of male and female subjects. Diagnosed COVID-19 patients admitted to the ward or ICU within 2 weeks from the onset of the disease onset were included in the study, whereas patients with any prior systemic disease were excluded from the study. Additionally, 20 healthy adults with negative COVID-19 PCR tests were selected as controls**.**
After collection, blood samples were centrifuged, and the serum was separated, labelled, and stored in aliquots at -80 °C [11]. All biochemical parameters were measured on an automated chemistry analyzer, AU 680 Beckman & Coulter and Sysmex [11,12]. Liver function tests: Alanine aminotransferase (ALT), Aspartate aminotransferase (AST), Alkaline phosphatase (ALP), Total and Direct Bilirubin), Renal function tests: Blood urea nitrogen (BUN), Creatinine, and cardiac Troponin I (Trop I) were performed to assess the organ function. Inflammatory markers, including C-reactive protein (CRP), interleukin-6 (IL-6), Lactate dehydrogenase (LDH), and Ferritin, were also measured to predict the extent of inflammation in these patients. In addition, serum electrolytes, including sodium (Na) and Potassium (K), as well as hematological markers such as Complete Blood Count and D-Dimer, were also measured. All parameters were compared and analyzed between the two genders across the three groups: ICU, Ward, and Control. Nasopharyngeal swabs of healthy controls were collected from the posterior nasopharynx and removed while rotating. Total RNA was extracted according to the guidelines for the qualitative detection of SARS-CoV-2 by real-time PCR [13].
Multinomial logistic regression analysis was then applied to predict the level of disease severity in normal individuals, as well as in patients with moderate and severe COVID-19. A dataset preprocessed in *Python* was utilized with the help of *Jupyter notebook*. The preprocessing steps included the removal of unnecessary fields, a categorical data tagging method, and splitting of the data sets into training (80%) and testing (20%) sets. A multinomial logistic regression model was used to handle multiclass classification problems, executed with the help of the LBFGS solver. Receiver Operating character (ROC) analysis was also performed in which the ROC curve provided a display of the model's performance in distinguishing between the three classes by plotting the true positive rate against the false positive rate. The area under the ROC curve (AUC) reflects the performance of the model, and for a perfect test, the AUC ranges between 1 and 0. The cross-validation technique is also applied to reduce overfitting and ensure that the model is not only accurate but also generalizable.

Statistical analysis was performed using the R programming language. After testing the normality of data, the parametric Student's t-test and non-parametric Mann-Whitney U-test were applied wherever required for gender-based comparison of all the biochemical markers between each group, i.e., controls, ward, and ICU. All the results were expressed as mean ± SD or median (Interquartile range IQR). The probability or $p\text{-value} < 0.05$ was considered to be statistically significant. The collective data is also presented in the form of box and whisker plots. Regression analysis was done using Python. The study employed data preprocessing, model training, and evaluation by employing key performance indicators, including precision, recall, and the F1 score. The importance of the variables in the multinomial regression analysis was shown through a variable importance graph, whereas the performance of the model was assessed and presented in the form of a confusion matrix and ROC/AUC analysis.

Results
Biochemical analysis
In the control group, no significant differences were observed with respect to age in either gender ($p = 0.226$), ward ($p = 0.787$), or ICU ($p = 0.201$) participants. The median age of controls, ward, and ICU subjects was 36, 54, and 53.5 years, respectively. Whereas the median age for both genders (male & female) in each group was control (41 & 30.5 years), ward (58.5 & 48.5 years), and ICU (51.5 & 54.5 years). Biochemical markers, including renal function tests (Blood urea nitrogen-BUN, Creatinine), liver function tests (Alanine aminotransferase-ALT, Aspartate aminotransferase-AST, Alkaline phosphatase-ALP, ALP, Total Bilirubin and Direct Bilirubin), inflammatory markers (C-reactive protein-CRP, Interleukin six-IL-6, Lactate dehydrogenase-LDH and Ferritin), hematological markers (White blood cells-WBCs, Neutrophils, Lymphocytes, Platelets, Hemoglobin-Hb and D-Dimer), cardiac marker (Troponin I-Trop I), and electrolytes (Sodium-Na and Potassium-K) were evaluated in all the participants of the study. The values of all biochemical markers in each of the three groups — Control, Ward, and ICU —were measured and compared based on gender. All the data is presented in the form of median (IQR) and mean ± SD.

Gender-based comparison in the Control group

The gender-based comparison of the Control group is shown in Table 1. The p-values of all markers were insignificant except for BUN, Hemoglobin, and Ferritin, which were 0.048, 0.0004, and 0.0005, respectively.

Table 1. Gender-based comparison of all biochemical markers in the Control Group given as median (Q1; Q3) and mean ± SD

	CONTROL GROUP			
		Males (n=10)	**Females (n=10)**	***P*-Value**
RENAL MARKERS	**Blood urea nitrogen (mg/dL)**	15 (12.3;17)	10 (8.3;12.5)	0.0483*
	Creatinine (mg/dL)	0.8 (0.7;0.8)	0.6 (0.5;0.7)	0.1139
HEPATIC MARKERS	**Total Bilirubin (mg/dL)**	0.55 (0.5;0.67)	0.65 (0.5;0.78)	0.8480
	Direct Bilirubin (mg/dL)	0.15 (0.1;0.2)	0.2 (0.1;0.2)	0.7382
	Alkaline phosphatase (U/L)	94 (88.75;101)	79.5 (74;95)	0.1858
	Alanine aminotransferase (U/L)	21.5 (18.3;28.3)	17.5 (15;22.5)	0.1852
	Aspartate aminotransferase (U/L)	26 (22.25;30.5)	23 (21.25;28.5)	0.6763
INFLAMMATORY MARKERS	**Lactate dehydrogenase (U/L)**	169.5 (154;183)	159.5 (148;173)	0.2413
	C-reactive protein (mg/dL)	0.65 (0.45;0.87)	0.62 (0.5;0.78)	0.6755
	Interleukin-6 (mg/dL)	2.8 (1.95;4.62)	2.6 (1.275;3.91)	0.4053
	Ferritin (ng/mL)	271 (224;351)	100 (72;121)	0.0004*
HEMATOLOGICAL MARKERS	**Hemoglobin (g/dL)**	13.7 (13.3;14.1)	11.9 (11.1;12.4)	0.0005*
	White blood cells ($x10^9$/L)	6 (5.6;7.3)	8.3 (6.7;8.9)	0.1208
	Neutrophil %	49.8 (47.7;51.5)	51.4 (49.7;54.4)	0.3074
	Lymphocyte %	31 (27.7;36)	37.6 (33.32;39)	0.1398
	Platelets ($x10^9$)	215.2 ± 63.2	236.9 ± 60.9	0.448
	D-Dimer (µg/mL)	0.33 (0.18;0.49)	0.61 (0.32;0.7)	0.0889
CARDIACMARKER	**Troponin-I (ng/mL)**	0.03 (0.02;0.04)	0.02 (0.01;0.03)	0.3287
ELECTROLYTES	**Sodium (mEq/L)**	139.8 ± 3.4	139.9 ± 2.8	0.9451
	Potassium (mEq/L)	4.14 ± 0.5	4.16 ± 0.4	0.9277

***p-value < 0.05**

Gender-based comparison in the Ward Group

The gender-based results and comparison of all biomarkers in moderate (ward) patients affected by SARS-CoV-2 are presented in Table 2 as median (IQR) and mean ± SD. Renal, hepatic, and cardiac markers showed no significant differences in both genders. It is interesting to note here that although most of the hematological markers, including WBC, Neutrophil, Lymphocyte, Platelets, and D-Dimer, along with inflammatory markers, including LDH, CRP, and Ferritin, show higher values in females than in males, but only CRP and LDH are statistically significant with *p*-values of 0.0138 and 0.0028, respectively.

Table 2: Gender-based comparison of all biochemical markers in the ward (moderate) COVID-19 patients expressed as median (Q1; Q3) and mean ± SD

	WARD			
		Males (n=10)	**Females (n=10)**	***P*-Value**
RENAL MARKERS	**Blood urea nitrogen (mg/dL)**	33 (16.5;45.8)	24.5 (19.5;31.8)	0.4495
	Creatinine (mg/dL)	1.35 (1.02;1.55)	0.95 (0.67;1.25)	0.1496
HEPATIC MARKERS	**Total Bilirubin (mg/dL)**	0.5 (0.3;0.7)	0.5 (0.4;0.6)	0.8190
	Direct Bilirubin(mg/dL)	0.2 (0.1;0.2)	0.1 (0.1;0.3)	0.3382
	Alkaline phosphatase (U/L)	97.5 (73.5;121)	96.5 (75.5;135)	0.8205
	Alanine aminotransferase (U/L)	27 (15.5;53.3)	26.5 (21.3;36.5)	0.8500
	Aspartate aminotransferase (U/L)	41.5 (33.3;73.3)	43 (25.5;58)	0.6499
INFLAMMATORY MARKERS	**Lactate dehydrogenase (U/L)**	325 (277.3;435.1)	659.3 (480;805)	0.0028*
	C-reactive protein (mg/dL)	17 (12;25.5)	46.8 (30.3;92.4)	0.0138*
	Interleukin-6 (mg/dL)	3.8 (0.98;5.75)	3.25 (1.17;4.85)	0.4958
	Ferritin (ng/mL)	541.5 (129;1072)	765 (496;1358)	0.1858
HEMATOLOGICAL MARKERS	**Hemoglobin (g/dL)**	10.9 (10.2;13.3)	9.6 (8.9;12.9)	0.3844
	White blood cells ($x10^9$/L)	10.75 (9.1;14.9)	8.7 (6.8;11.5)	0.2730
	Neutrophil %	87.9 (82.8;90.02)	82.3 (79.4;86.1)	0.3846
	Lymphocyte %	6.35 (3.5;8.22)	10.05 (5.8;13.7)	0.2411
	Platelets ($x10^9$/L)	185 ± 120.1	283 ± 168.0	0.1526
	D-Dimer (µg/mL)	156 (31.2;179.5)	303 (150;875)	0.1212
CARDIAC MARKER	**Troponin-I (ng/mL)**	0.006 (0.001;0.03)	0.01 (0.001;0.02)	0.9081
ELECTROLYTES	**Sodium (mEq/L)**	134.7 ± 6.5	135.4 ± 5.6	0.8017
	Potassium (mEq/L)	3.98 ± 1.0	4.42 ± 0.9	0.3494

***p-value < 0.05**

Gender-based comparison in the ICU group

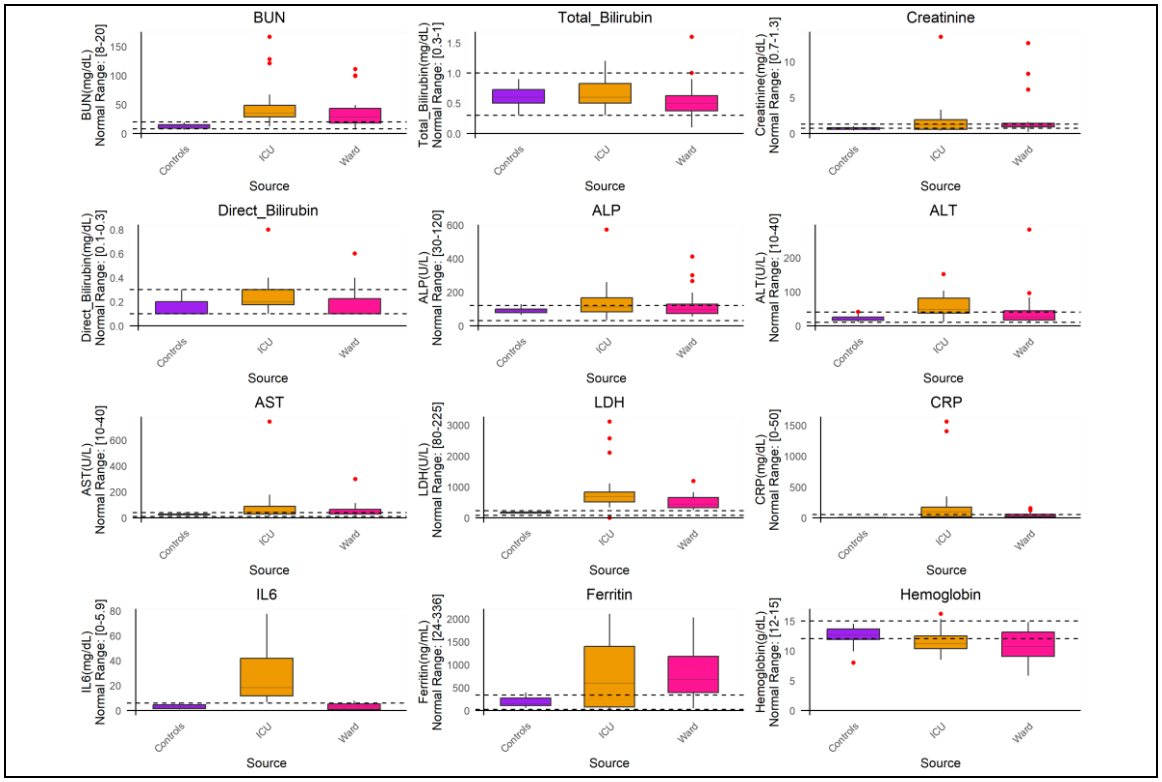
Table 3 gives the gender-based comparison of severe (ICU) patients of COVID-19. Similar to the ward results, hematological markers, including WBC, Lymphocytes, Platelets, and D-Dimer, show a relative increase in females compared to males. Whereas, in the inflammatory markers, except LDH, all the other ones, namely Ferritin, IL-6, and CRP, showed much lesser values in females compared to males, with CRP being significantly higher in males than female ICU patients (*p*-value = 0.0309). Moreover, unlike the ward patients, LDH was not significantly different (p-value = 0.3846) between the genders in this group.

Table 3: Gender-based comparison of all biochemical markers in ICU (severe) COVID-19 patients expressed as median (Q1; Q3) and mean ± SD

	ICU			
		Males (n=10)	**Females (n=10)**	***P*-Value**
RENAL MARKERS	**Blood urea nitrogen (mg/dL)**	41 (27.3;62.5)	34 (29.5;40.8)	0.5703
	Creatinine (mg/dL)	1.3 (0.8;2.72)	1.1 (0.52;1.55)	0.4250
HEPATIC MARKERS	**Total Bilirubin (mg/dL)**	0.6 (0.5;0.87)	0.6 (0.42;0.77)	0.4443
	Direct Bilirubin (mg/dL)	0.2 (0.2;0.27)	0.2 (0.12;0.27)	0.6888
	Alkaline phosphatase (U/L)	113 (92;149)	96 (78;182)	0.5707
	Alanine aminotransferase (U/L)	48.5 (38.8;81.5)	49.5 (35.5;75.8)	0.8797
	Aspartate aminotransferase (U/L)	63 (29;88.8)	46.5 (33.5;73.8)	0.8498
INFLAMMATORY MARKERS	**Lactate dehydrogenase (U/L)**	688.5 (474;711)	713(558;1791)	0.3846
	C-reactive protein (mg/dL)	191.9 (115;334)	30.5 (13.7;83.3)	0.0309*
	Interleukin-6 (mg/dL)	24 (13.6;40.3)	13.5 (8.5;39.1)	0.2413
	Ferritin (ng/mL)	860.5 (84.6;1565)	444.1 (88;1223)	0.9698
HEMATOLOGICAL MARKERS	**Hemoglobin (g/dL)**	11.6 (10.5;13.8)	11 (10.3;11.6)	0.5704
	White blood cells (x10^9/L)	16.4 (12.5;17.1)	19.3 (11.3;34.1)	0.3074
	Neutrophil %	90.4 (88.5;92.6)	87.9 (83.5;92.9)	0.7336
	Lymphocyte %	2.9 (2.2;9.5)	5.5 (3.9;9.6)	0.3255
	Platelets (x10^9/L)	203.7 ± 136.9	240.7 ± 119.5	0.5282
	D-Dimer (µg/mL)	4.95 (2.7;1079)	10.96 (6.2;462)	0.7337
CARDIAC MARKER	**Troponin-I (ng/mL)**	0.04 (0.02;0.06)	0.12 (0.01;0.18)	0.5701
ELECTROLYTES	**Sodium (mEq/L)**	136.5 ± 8.7	139.3 ± 6.0	0.4176
	Potassium (mEq/L)	4.18 ± 0.7	4.39 ± 0.6	0.5336

***p*-value < 0.05**

The data for all biochemical markers in male and female subjects of this study can also be visualized in the form of box and whisker plots, as presented in Figure 1. The data in these plots is shown in three different colors, denoting the controls, ward, and ICU groups. These plots facilitate the comparison of data distribution across all study groups.



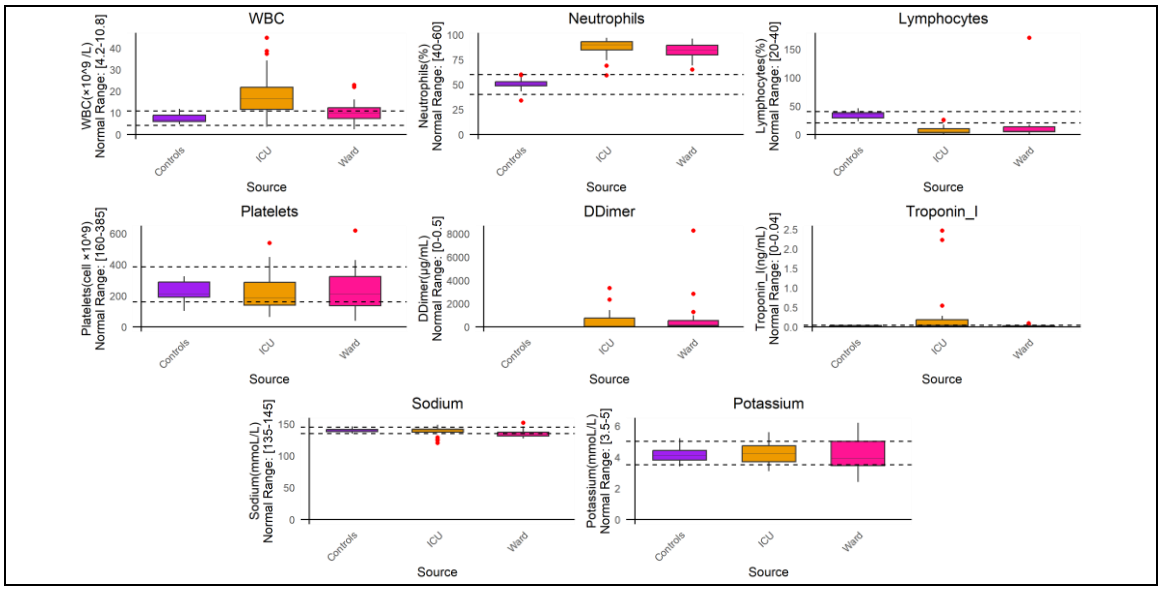


Figure 1: Box and whisker plots showing data spread of each biomarker in controls, ward, and ICU patients, denoted by purple, pink, and orange colors, respectively. The height of the box represents the interquartile range.

Multinomial Regression Analysis

Performance indicators from the multinomial regression analysis for all study groups are presented in Figure 2.

Performance Metrics:

Normal Class: The model performed exceptionally, with precision, recall, and F1 score all equal to 1. It points to the fact that it performs very well in predicting the normal class.

Moderate Class: In this class, the precision was 0.60, the recall was 0.75, and the F1-score was found to be 0.67. Therefore, it can be concluded that the model achieved moderate success in recognizing moderate cases.

Severe Class: This class exhibited reduced accuracy, with precision recorded at 0.67, recall at 0.50, and an F1-score of 0.57. The model struggles to identify severe cases, demonstrating only 57% accuracy.

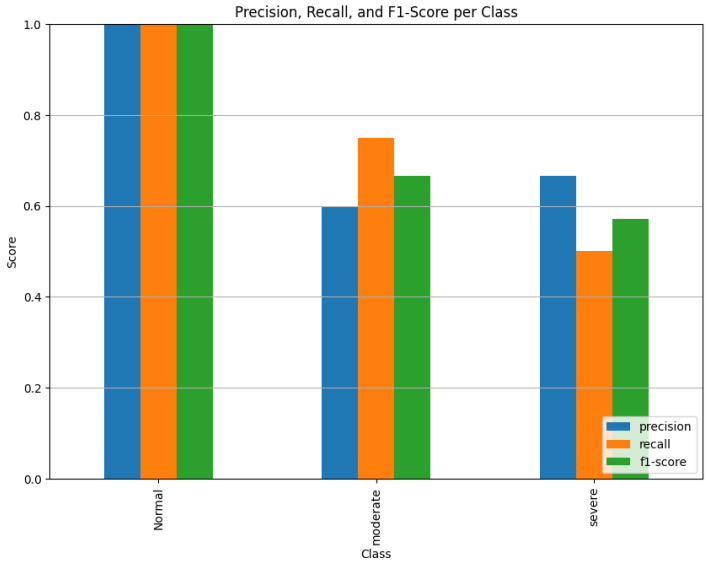


Figure 2: Graphical representation of performance metrics in all groups of the regression model

Confusion Matrix and Misclassification:

From the confusion matrix (shown in Figure 3), it was observed that the model had greater accuracy in deciding normal cases without any misclassification. Yet, it moderately misclassified the cases in the other two severity levels, both in terms of assigning severe severity to moderate ones and vice versa.

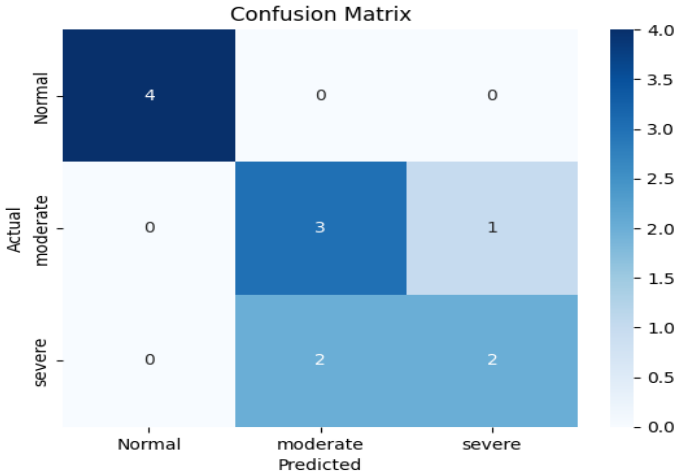


Figure 3: Confusion Matrix

ROC and AUC Analysis:

The receiver operating characteristic (ROC) curve analysis (Figure 4) showed that the model's ability to differentiate between patients in the normal group was very high, with an AUC value of 1.0. Nonetheless, the AUC of the classifier for the moderate and severe classes was both 0.81. The model showed excellent performance in predicting normal patients, while it showed less capability in classifying moderate and severe diseased patients.

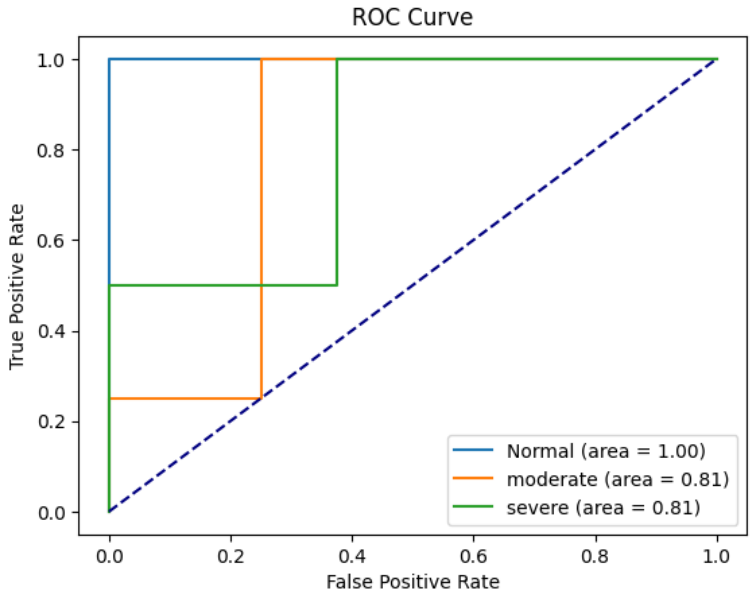


Figure 4: Receiver Operating Characteristic curve

Cross-Validation Results:

Cross-validation is used to validate the model's performance. For this purpose, data is trained over different subsets or folds. The model is then tested on different portions of the data. This step helped detect the model's overfitting and ensure it generalizes well across the data. 5-fold cross-validation is performed and evaluated by building the ROC curve, which plotted the True Positive Rate against the False Positive Rate at various threshold settings. As seen in Figure 5. ROC curves display the AUC values at different fold training, ranging from 0.81 to 0.94. The AUC values of the 0-fold, 1-fold, 2-fold, 3-fold, and 4-fold CV models are displayed as 0.94, 1.00, 0.88, 0.81, and 0.91, respectively. A High AUC value showed the excellent discriminative ability for that particular fold. The Blue curve in the graph displays the mean AUC value of 0.90 with a standard deviation of 0.06. An AUC value close to 1.00 indicates the model's consistent performance across different subsets of the data.

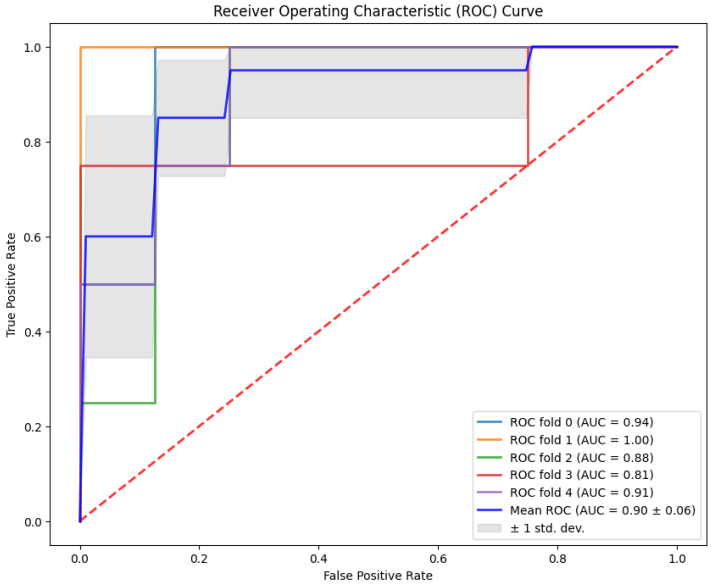


Figure 5: ROC curve obtained after 5-fold cross validation

Variable Importance Graph:

This was built to identify the most influential biochemical variables for this model. Figure 6 shows that AST, IL-6, Direct bilirubin, and LDH have the most significant impact on classifying the disease status among the three groups. Elevated levels of these biomarkers might correlate with the severity of COVID-19, whereas their lower levels may indicate a normal or moderate degree of the disease. Moving down the graph, it is evident that Neutrophils, ALT, and CRP may also play important roles in defining disease status, but they are less significant than the topmost indicators described above. Age, sex, total bilirubin, and other biomarkers, displayed along the axis with lower scores, suggest that these markers have a lesser impact on the model's ability to classify subjects as normal, moderately affected, or severely affected with COVID-19.

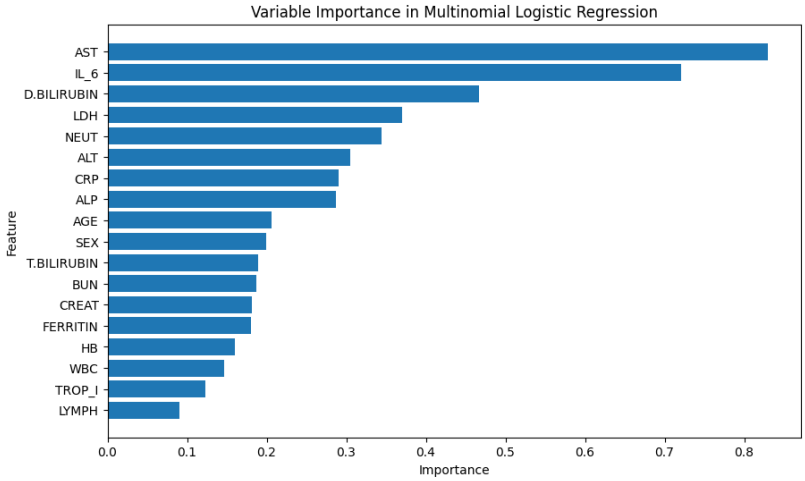


Figure 6: Variable importance graph of multinomial regression model

Discussion

The clinical presentation of SARS-CoV-2 infection ranges from patients being totally asymptomatic to patients with mild, moderate, severe, and critical forms of the disease [14]. Early stratification of these patients based on their potential to develop a mild, moderate or severe disease will ensure effective utilization of our limited health resources and reduce the fatality indices associated with this pandemic [8]. Predictive biochemical parameters can serve as a backbone in achieving this goal. Therefore, we evaluated a number of biomarkers in the three study groups, which were classified based on disease intensity as normal, moderate, and severe. Our results showed a progressive increase in the levels of these biomarkers, which was directly proportional to the severity of the disease.

The biochemical evaluation based on gender was performed within each group. The results revealed increased levels of most hematological and inflammatory markers in females of both disease groups as compared to males. Specifically, significantly higher levels of CRP and LDH were observed in females compared to males in the moderately diseased group. On the other hand, in the severely affected group, LDH levels were elevated in females more than in males; however, CRP levels were significantly elevated in males rather than females. This observation aligns with another study, which states that men severely affected by SARS-CoV-2 have higher CRP levels than females, independent of age and other comorbidities [15]. Thus, our findings suggest that LDH may be an effective predictor of disease progression in females, whereas higher CRP levels are

indicative of severe disease in males affected by SARS-CoV-2 infection. Although another gender-based biochemical analysis with a greater sample size would help to further validate these findings.

A key feature that determines the severity of SARS-CoV-2 infection is inflammation. CRP and LDH are established biomarkers of inflammation and tissue damage, which are found to be markedly elevated after exposure to this virus [14]. Numerous studies in recent years have reported their relevance in the progression of COVID-19, but hardly any have evaluated the gender differences in relation to the intensity of the disease.

The multinomial regression predictive model was built to classify COVID-19 patients according to their disease status as normal, moderately, and severely affected [16]. This model demonstrated a high ability in classifying normal patients but encountered challenges in differentiating moderate and severe patients. The model achieved an accuracy of 75% with excellent performance in predicting normal cases. The performance of the model was evaluated using precision, recall, and F1 score values (Fig. 2). As a cumulative performance indicator, the F1 score was 1.0 for normal cases and 0.67 and 0.57 for moderate and severe cases, respectively. These findings were consistent with a similar study conducted in Korea, which also highlights the difficulty of accurately predicting severe cases of COVID-19 [4]. Despite this, overall performance, as validated by cross-validation, displayed a mean AUC value of 0.91 across different folds. The mean, high AUC value of the model indicates its strong classifying capacity. The cross-validation of the model reduced the chances of overfitting and ensured robustness when tested on unseen data.

The variable importance analysis (Fig.6) revealed that AST, IL-6, Direct Bilirubin, and LDH were the most influential factors in predicting disease severity, particularly for severe cases. Whereas, Albumin, Lymphocyte %, and Prothrombin time were identified as major predictors of disease severity in another study [4]. The consistency in elevated levels of AST in both studies confirms the importance of this biomarker as a determinant of disease progression in COVID-19. Other variables along the graph, such as Neutrophils, ALT, and CRP, contributed moderately to the classification process, while factors like Age, Sex, and Total Bilirubin had a lesser impact on the model's performance.

A larger and more diverse sample collection can further enhance the robustness of this model and minimize any biases that may be present. Additionally, the integration of this predictive model into the clinical decision support system enables real-time assessment that may aid in making effective and timely decisions.

Conclusion

A statistical comparison of various biochemical markers in males and females of normal, moderate, and severe COVID-19 patients revealed LDH as the most important predictor of disease severity in females, whereas in males, CRP emerged as the most important marker of severity. A multinomial regression model with predictive ability to distinguish between the three given study groups displayed an AUC value of 1.00 for the Normal group and 0.81 for both Moderate and Severe groups. To mitigate overfitting and poor performance on unseen data, cross-validation was performed using a 5-fold training approach, yielding a mean AUC value of 0.91 ± 0.06, indicating that the model exhibits excellent discriminative ability for disease status in COVID-19 patients. AST, IL-6, direct bilirubin, and LDH emerged as the major predictors of disease progression in this viral infection.

Author Contributions: Visualization, investigation, sample collection, formal analysis, and Writing - original draft M.K.; Supervision, analysis, and review and editing A.A., Z.H., T.M.; methodology, software, and writing—review and editing H.M.H., M.A.T;

Conflicts of Interest: The authors declare no conflicts of interest

References

1. Gacche RN, Gacche RA, Chen J, Li H, Li G. Predictors of Morbidity and Mortality in COVID-19. Eur Rev Med Pharmacol Sci. 2021;25(3):1684–1704. https://doi.org/10.26355/eurrev_202102_24815
2. Alimohamadi Y, Sepandi M, Taghdir M, Hosamirudsari H. Determine the Most Common Clinical Symptoms in COVID-19 Patients: A Systematic Review and Meta-analysis. J Prev Med Hyg. 2020;61(3):E304–E312. https://doi.org/10.15167/2421-4248/jpmh2020.61.3.1530
3. Hashmi HA, Asif HM. Early Detection and Assessment of COVID-19. Front Med (Lausanne). 2020;7:311. https://doi.org/10.3389/fmed.2020.00311
4. Yoo EH, Chang SH, Song DY, et al. Comprehensive Laboratory Data Analysis to Predict the Clinical Severity of Coronavirus Disease 2019 in 1,952 Patients in Daegu, Korea. Ann Lab Med. 2022;42(1):24–35. https://doi.org/10.3343/alm.2022.42.1.24
5. Gallo Marin B, Aghagoli G, Lavine K, Yang L, Siff EJ, Chiang SS, et al. Predictors of COVID-19 Severity: A Literature Review. Rev Med Virol. 2021;31(1):e2146. https://doi.org/10.1002/rmv.2146
6. Cecconi M, Piovani D, Brunetta E, Aghemo A, Greco M, Ciccarelli M, et al. Early Predictors of Clinical Deterioration in a Cohort of 239 patients Hospitalized for Covid-19 Infection in Lombardy, Italy. J Clin Med. 2020;9(5):1548. https://doi.org/10.3390/jcm9051548
7. Qi J, He D, Yang D, Wang M, Ma W, Cui H, et al. Severity-Associated Markers and Assessment Model for Predicting the Severity of COVID-19: A Retrospective Study in Hangzhou, China. BMC Infect Dis. 2021;21(1):774. https://doi.org/10.1186/s12879-021-06511-7
8. Sarhan AR, Hussein TA, Flaih MH, Hussein KR. A Biochemical Analysis of Patients With COVID-19 Infection. Biochem Res Int. 2021;2021:1383830. https://doi.org/10.1155/2021/1383830

9. Akdogan D, Guzel M, Tosun D, Akpinar O. Diagnostic and early prognostic value of serum CRP and LDH levels in patients with possible COVID-19 at the first admission. J Infect Dev Ctries. 2021 Jun 30;15(06):766-72.
10. Shiri I, Mostafaei S, Haddadi Avval A, Salimi Y, Sanaat A, Akhavanallaf A, et al. High-dimensional multinomial multiclass severity scoring of COVID-19 pneumonia using CT radiomics features and machine learning algorithms. Sci Rep. 2022 Sep 1;12(1):14817.
11. Kennedy AD, Ford L, Wittmann B, Conner J, Wulff J, Mitchell M, et al. Global biochemical analysis of plasma, serum and whole blood collected using various anticoagulant additives. PLoS One. 2021;16(4):e0249797. http://dx.doi.org/10.1371/journal.pone.0249797
12. Farid E, Sridharan K, Alsegai OA, Khawaja SA, Mansoor EJ, Teraifi NA, et al. Utility of inflammatory biomarkers in patients with COVID-19 infections: Bahrain experience. Biomark Med. 2021;15(8):541–9. http://dx.doi.org/10.2217/bmm-2020-0422
13. Iwasaki S, Fujisawa S, Nakakubo S, Kamada K, Yamashita Y, Fukumoto T, et al. Comparison of SARS-CoV-2 detection in nasopharyngeal swab and saliva. J Infect. 2020;81(2):e145–7. http://dx.doi.org/10.1016/j.jinf.2020.05.071
14. Hassan SA, Sheikh FN, Jamal S, Ezeh JK, Akhtar A. Coronavirus (COVID-19): a review of clinical features, diagnosis, and treatment. Cureus. 2020 Mar;12(3).
15. Qin L, Li X, Shi J, Yu M, Wang K, Tao Y, Zhou Y, Zhou M, Xu S, Wu B, Yang Z. Gendered effects on inflammation reaction and outcome of COVID-19 patients in Wuhan. J Med Virol. 2020 Nov;92(11):2684-92.
16. Mphekgwana PM, Sono-Setati ME, Mokgophi TV, Kifle YG, Madiba S, Modjadji P. Retrospective Analysis of the Outcome of Hospitalized COVID-19 Patients with Coexisting Metabolic Syndrome and HIV Using Multinomial Logistic Regression. Int J Environ Res Public Health. 2023 May 12;20(10):5799