



Uncovering COVID-19 Death Risk for Life on the Line with Machine Learning Precision

Muhammad Adnan Hasnain^{1,*}, Tahir Abbas², Jamshaid Iqbal Janjua³ and Sunawar Khan⁴

¹Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila, Italy

²Department of Computer Science, The TIMES Institute, Multan 60000, Pakistan

³Al-Khawarizmi Institute of Computer Science (KICS), University of Engineering & Technology (UET), Lahore, Pakistan

⁴International Islamic University, Islamabad 44000, Pakistan

Abstract

The unprecedented pressure on healthcare systems globally because of the unprecedented COVID-19 pandemic means there is an urgent need for innovations that can be used to guide critical decision-making. Within this study, we apply machine learning to predict the risk of mortality in COVID-19 using a dataset of over a million records. The results of the analysis are used to explore the contribution of demographics, symptoms, and preexisting conditions to the predictions with the help of advanced models like Decision Tree, Random Forest, Logistic Regression, Support Vector Machine, Gradient Boost Classifier, and a very high-performing Voting Classifier. With some feature engineering and some data preprocessing, we were able to get some excellent and valuable insights out of the data, and the Voting Classifier is capable of hitting a whopping 93% accuracy, while both Random Forest and Logistic Regression

always hit 91% within the test data. Results show that the significant risk factors are Age and preexisting disease or health condition, and subtle and variable risk patterns are uncovered by symptoms and socioeconomic and demographic profiles. In addition, the model performance was validated by the F1-score and ROC curves to be learned from and robust. This study also shows how applying machine learning techniques can drive actionable insights for public health interventions beyond 'predictive analytics'. Our approach allows healthcare providers to allocate resources intelligently, increase the use of resources among high-risk patients, prioritize and deliver care, and eventually improve survival rates. Beyond the advancements in understanding COVID-19 mortality, this research shows the critical importance of machine learning to help solve worldwide health crises and lays the foundation of data-driven approaches for handling future problems.

Keywords: COVID-19, health care analytics, predictive modeling mortality risk.

1 Introduction

The COVID-19 pandemic has unquestionably stressed the global healthcare systems, showcasing the urgency



Academic Editor:

Abdur Rehman Sakhawat

Submitted: 18 March 2025

Accepted: 22 March 2025

Published: 30 March 2025

Vol. 1, No. 1, 2025.

10.62762/TNC.2025.507897

*Corresponding author:

✉ Muhammad Adnan Hasnain

muhammadadnan.hasnain@graduate.univaq.it

Citation

Hasnain, M. A., Abbas, T., Janjua, J. I., & Khan, S. (2025). Uncovering COVID-19 Death Risk for Life on the Line with Machine Learning Precision. *ICCK Transactions on Neural Computing*, 1(1), 30–43.



© 2025 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

for innovative methods for improving patient care and resource management. Amidst the early days of the COVID-19 pandemic, we've learned a lot, but there's still a lot we don't know about risk factors for mortality. To do so, this work uses machine learning to predict the mortality risk from COVID-19. This research provides actionable insights that allow clinicians to make better decisions leading to improved survival outcomes, using demographic characteristics, symptoms, and preexisting medical conditions.

The COVID-19 pandemic has overwhelmed healthcare systems, and millions of people have died. Understanding COVID-19 mortality risk variables is critical for improving patient outcomes and allocating related resources. Many of the complicated demographics, symptoms, and preexisting illnesses that make people susceptible to dying from COVID-19 are not going to be captured by traditional risk assessment studies. Finally, analyzing these parameters using machine learning may result in identifying high-risk individuals and the ability to perform targeted therapy. COVID-19 is caused by Coronavirus type 2. According to the WHO, since it first emerged in Wuhan, China, in December 2019, this global epidemic has infected nearly 254 million people since November 16, 2021. Two years back, when COVID-19 broke out, the most significant worldwide health issue. We do not have a cure for this pandemic, and we know nothing about its dynamics, so it is very severe. Coronaviruses can occur by causing colds, MERS, and SARS. On March 30, 2020, the WHO reported a 4.6% death rate for certified COVID-19 infections. The percentage was Age, immunological status, and region-dependent, varying from 0.2 to 15% [1].

Global healthcare systems suffer from COVID-19. Non-pharmaceutical treatments are crucial national policies because, despite aggressive immunization programs, many nations will not have universal vaccine access until 2023 [2].

17–45% of COVID-19 cases are symptomless and do not require therapy [3, 4]. On the contrary, the global case death rate ranges from 2–3% [5]. Fever, persistent cough, anosmia, and dyspnea are common symptoms between these extremes and can be managed at home or impatiently. Limiting the virus's impact requires knowing which people are most susceptible to severe sickness and need the most significant resources.

The risk of severe disease must be considered in all decisions. Certain patient traits, comorbidities, and

lifestyle variables increase the chance of mortality or severe illness after infection [6, 7]. Once SARS-CoV-2 is contracted, symptoms and vital signs can help predict prognosis [8]. Laboratory testing and imaging can stratify risk for early, intensive management, but only hospital in patients with this data, who are often already severely affected [9, 10].

Risk classification and population management at scale are supported by robust, predictive models for COVID-19 acquisition and prognosis [11] and resource management [12]. These models inform organizational decision-making. A major COVID-19 risk model, COVID, uses primary care data from 8 million adults and has external validation¹⁹. The NHS uses it for clinical assessment²⁰. However, the individual is disregarded, and granular, patient-specific risk-scoring may unify decision-making at all levels. However, current individualized risk ratings typically confuse COVID-19 acquisition risk with infection mortality risk [13], limiting their use in patient treatment.

Risk factor [14] prediction models will only be effective if they are, for instance, cheap and easily accessible to the general public without requiring specialized testing or trips to the hospital for evaluation. Using risk prediction techniques, patient triage should improve efficiency and confidence in reducing hospital burden with hospital-at-home solutions during the pandemic. Risk scores should be dynamic and contemporaneous and include symptoms and vital sign data to benefit clinical and research teams. The main goal of this study is to use a large, rich dataset and a selected, clinically informed approach to create and verify a population-based prediction model for the dynamic prediction of COVID-19 mortality risk for confirmed diagnosis.

The sheer scale of COVID-19 has been both eye-opening and has laid bare some critical weaknesses in conventional risk assessment and resource distribution tools, which have not been adaptive enough to the fast-evolving and converging risks we identify in this piece. The issue is that around the globe, healthcare systems are sorely pressed to keep care robust yet affordable, with limited resources available to them. Machine learning brings strong potential to these challenges with its ability to process big datasets and find these patterns. This study is motivated by the need to utilize advanced analytics for proper patient prioritization, leading to a better survival rate and less burden on a healthcare provider.

This research fills a gap between data-driven insights and their application to further improve global health responses to the pandemic. This research aims to design a machine learning-based classification model that reliably and accurately predicts COVID-19 mortality with high precision.

This study aims to create models to identify the high and low-risk groups of patients by examining critical factors like Age, symptoms, and prior medical history, which will give healthcare professionals additional decision-making options by taking advantage of data-driven tools. Such predictive tools are crucial for identifying vulnerable individuals early on so that they get prioritized care. Moreover, this research aims to optimally share scarce healthcare resources (ICU beds, ventilators, and personnel) with those who need to be reached. These findings should inform immediate clinical applications and public health policy that identify demographic and symptomatic trends that lead to excess mortality risk. Lastly, this work seeks to reduce the overall burden of the pandemic by increasing overall survival, optimizing healthcare systems, and ultimately gaining a better understanding of the determinants that drive COVID-19 severity.

The paper introduces a machine learning framework that combines Decision Trees, Random Forests, Logistic Regression, Support Vector Machines, and Gradient Boost Classifiers to predict COVID-19 mortality risk. Uses advanced preprocessing and feature engineering techniques to maximize predictive accuracy. The Voting Classifier produces 93% Predictive Accuracy; Random Forest and Logistic Regression consistently achieve 91% testing accuracy. It finds that Age and preexisting conditions are the leading mortality risk factors, yet it reveals nuanced patterns between symptoms and demographic variables. Generates actionable insights for healthcare providers to identify high-risk patients and to develop measures for wide-scale public health interventions.

The study has been presented systematically, where the research problem is investigated and solutions proffered systematically. In the Introduction, we present the problem, the motivations, the high-level objectives, and the contributions. The Literature Review next depicts prior work on COVID-19 risk prediction and identifies research gaps. In Methodology, the works mentioned are the dataset, preprocessing, exploratory data analysis, machine learning models, and evaluation metrics (accuracy, F1-score, and ROC AUC). Results and

Discussion include key findings, a comparison of model performance, and implications. Finally, the Conclusion and Future Work section summarizes the findings, research hints, and paths for future work, such as working with other data features or developing hybrid modeling.

2 Existing Literature

The COVID-19 pandemic has accelerated research to discover factors associated with increased risk of dying and develop prediction tools to aid clinical and public health decision-making. This review synthesizes current work concerning COVID-19 risk prediction, risks, machine learning applications applied, and gaps in the literature that we have studied in this work.

The studies that have repeatedly found factors that raise the risk of a person catching COVID-19 are indeed diverse. The most important demographic factor was found to be Age, with older age groups finding a higher fatality risk than the expected proportion due to poorer immune response and greater prevalence of comorbidity [15]. Redondo et al. [16] have shown that people above 60 years are much more likely to face severe outcomes. They also have found that 'gender' is a risk factor in that male patients have higher mortality than female patients with reasons which may be a difference in their immune responses to reach or behaviors, such as smoking. In addition, diabetes, hypertension, cardiovascular diseases, chronic kidney disease, and obesity are also described as frequently as essential contributors to mortality. One example demonstrates that patients with many other conditions are exponentially more likely to have complications [17]. Further, there are associations between disease severity and specific symptoms like shortness of breath, fever, and fatigue. The studies, however, can only study each factor with a broad focus to see the interactions without being able to make a holistic risk profile.

It brought machine learning into COVID-19 research to make impossible analysis of large datasets and digging for non-linear relationships in the variable series. Traditional statistical models like logistic regression, used in some early models [18], were used to identify predictors of mortality. These methods were effective for simple relationships but proved effective for high dimensions and variable interactions.

Evaluation metrics to evaluate the predictive performance of various models must be robust. Popular metrics such as accuracy are insufficient if we have to evaluate a model with an imbalanced

dataset — look at the COVID-19 mortality prediction spread. However, F1-score, precision, Recall, and ROC AUC provide a more balanced estimation considering false positives and negatives. A study from Jose et al. [19] also shows the importance of precision-recall tradeoffs and how misclassification of high-risk patients in healthcare allows for catastrophic consequences. Machine learning techniques [20] have also been effectively applied to classify lung diseases, showcasing the potential of these methods in medical diagnostics. However, few studies have systematically compared models in terms of these metrics. For example, Rainio et al. [21] just focused on accuracy and did not consider the potential implications when choosing a value above all other things. This is problematic and underscores the requirement for a comprehensive model evaluation framework applicable to any reliability over any clinical scenario.

However, the literature in this area has still been incomplete. These studies demonstrate that first, limited feature integration is observed, as most of these studies analyze isolated variables (i.e., comorbidities and Age) instead of integrating demographics, symptoms, and comorbidities into a unified prediction model. Secondly, there are issues with poorly selected and executed preprocessing steps since models do not utilize advanced feature engineering and preprocessing steps proficiently, especially on the datasets that scale to the point where missing and erroneous data render the efficiency of the models susceptible. Next, although ensemble techniques have shown promise, the relative benefits of such techniques compared to single models have not been thoroughly investigated in COVID-19 risk prediction studies. Fourth, the evaluation metrics are usually very narrow and depend mainly on accuracy, while Recall, F1-score, and ROC-AUC can be important for such use cases as healthcare. Finally, insufficient consideration of real-world application is observed, where current studies only investigate a fraction of practical challenges in employing machine learning models in clinical or public health environments and constrain their translation potential.

The gaps in the literature on integrating demographic and other features of the patient with their gout symptoms are addressed in this study by combining demographic, symptom, and comorbidity features into one machine-learning model. Features that are significantly scaled and encoded for advanced preprocessing techniques to improve model performance and enable additional model robustness.

Several machine learning algorithms are evaluated using a comparative approach, including Decision Tree, Random Forest, Logistic Regression, Support Vector Machine, Gradient Boost Classifier, and Voting Classifier. Evaluation metrics are balanced with comprehensive evaluation metrics, accuracy, F1-score, Recall, precision, and ROC-AUC. In addition, this study improves the relationship between predictive analytics and very localized, real-world healthcare by providing insights on how to best prioritize patients and use existing resources.

3 Methodology

In the subsequent section, we present a comprehensive methodology for predicting COVID-19 mortality risk through machine learning. It refers to how we capture data, exploratory data analysis (EDA), data preprocessing, feature engineering, model selection and training, evaluation, visualization, etc. Figure 1 shows the research design workflow.

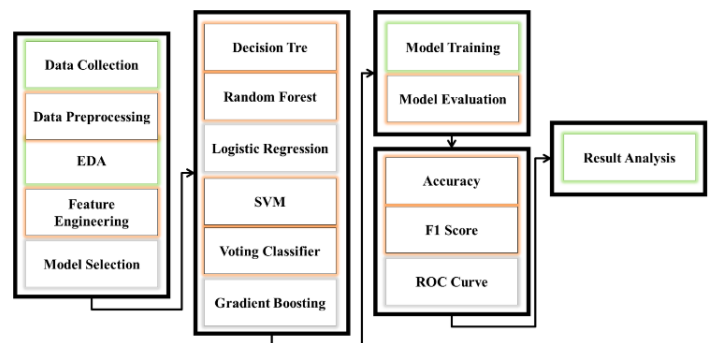


Figure 1. Research Design Workflow.

3.1 Data Loading and Exploration

We first looked at the dataset, which contained more than a million records, to check if the records were coherent and complete. This first exploration taught us something about the structure and quality of the data. Features with missing values were first examined in essential features like INTUBED and ICU, and features with lots of missing values were excluded since they were not helpful features. I changed the DATE_DIED column into a binary variable DIED to make an observable target for prediction where 1 is for deceased patients and 0 for survivors. Then, we scanned through the data structure to find AGE, SEX, preexisting conditions, or other key demographic and health-related features to use as modeling features.

3.2 Data Preprocessing

Before training an effective model on the dataset, it was preprocessed to make it clean. Because

97, 98, and 99 in the PREGNANT feature did not make sense in a numeric context, and so were replaced with NaN values. One hot encoding was used to transform categorical variables such as MEDICAL_UNIT and CLASIFFICATION_FINAL so machine learning algorithms could use them. StandardScaler and RobustScaler were applied to continuous variables (i.e., AGE), where the range of values was scaled, reducing the significance of large-magnitude features on model performance.

For feature selection and engineering, patterns, anomalies, and trends in the dataset were discovered with the help of Exploratory Data analysis (EDA). It was observed that the class distribution of the target variable DIED is highly imbalanced as far fewer patient was classified as deceased. This result indicated an unbalanced dataset, i.e., a need to train with a label balance dataset. We perform correlation analysis which shows strong relationships between predictors (comorbidities and AGE) and the target variable, and eventually help us choose features. Furthermore, mortality across different subgroups, such as age ranges, genders, and preexisting conditions, was assessed. We visualized these findings using techniques such as correlation heatmaps and feature heatmaps.

During the analysis, several visualizations were utilized to boost interpretability. Continuous variables like AGE were shown as histograms, and bar charts for categorical variables, i.e., SEX and DIABETES, were used to show the distribution and effect on the target variable. A correlation heatmap has been created to visualize relationships between numerical features, and important modeling predictors have been highlighted. A feature importance bar chart from the Random Forest model was used to present the most important predictors in the Feature Engineering phase. Lastly, in the Evaluation phase, confusion matrices were detailed to include true positives, false positives, true negatives, and false negatives for each model, and a comparative ROC curve was used to compare how the models balanced the sensitivity and specificity.

A feature correlation heatmap was also created in EDA to inspect variable relationships. In particular, the heatmap revealed that advanced Age and respiratory distress significantly increased the mortality of COVID-19, indicating that these were the major predictors. Moreover, it identified redundant features, which assisted in a more accurate final feature selection process, increasing predictive accuracy.

The Methodology incorporated these visualizations and analyses, providing easy-to-interpret and actionable insights that guided the study.

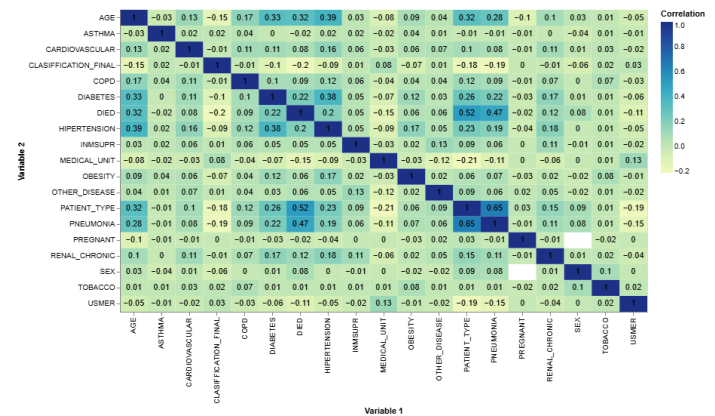


Figure 2. Feature Correlation Heatmap.

3.3 Feature Engineering

Feature engineering has proposed creating new features and refining existing ones to improve the dataset's predictive power optimally. Interaction terms between variables were created through polynomial features, which captured non-linear relationships. Feature selection was used to control for predictors with slight variance or not talking much about the target variable. In addition, we defined synthetic features (e.g., age buckets, count of comorbidities) to express these key risk factors in a way that is easy to interpret. In addition to improving the performance of these models, these transformations yielded a richer understanding of the inherent relationships.

3.4 Class Imbalance Handling

The dataset has a high imbalance between the number of records of patients who died and patients who survived, and it was found. To deal with this, I applied a random under-sampling technique to balance the dataset by selecting fewer samples of that majority class. We did this so that the models were trained over a balanced dataset to minimize the bias and increase the model's capacity to generalize to real-world scenarios. The model's performance was tested using the original and imbalanced dataset to check its applicability.

3.5 Model Selection and Training

- **Decision Tree Classifier:** Provided an interpretable baseline with a training accuracy of 95.7% and testing accuracy of 89.7%.
- **Random Forest Classifier:** An ensemble method combining multiple decision trees, achieving 91%

testing accuracy with intense precision and Recall.

- **Logistic Regression:** A simple yet effective model for binary classification, delivering an F1-score of 0.91 on both training and testing datasets.
- **Support Vector Machine (SVM):** Captured complex decision boundaries, particularly effective after feature scaling.
- **Gradient Boosting Classifier:** Utilized boosting techniques for high precision, Recall, and general robustness.

3.6 Evaluation Metrics

A suite of metrics was used to evaluate the performance of each model to ensure a balanced assessment. As an essential correctness measure, we use accuracy. Yet, precision, Recall, and F1 score are even more critical since minimizing false positives and negatives is of utmost importance in healthcare. First, we used the ROC-AUC metric to compare the models' discriminating ability across thresholds to differentiate between positive and negative cases. Confusion matrices were used to give insights into the distribution of true positives, false positives, and false negatives. These evaluations were applied to the models to ensure that the models do not only predict accurately but also are reliable predictions for real-world applications.

3.7 Interpretation and Insights

Based on key insights from analysis, it is found that Age and comorbidities are good predictors of survival of COVID-19. Feature importance plots, created from ensemble methods, such as Random Forest, indicated that Age was always the most significant predictor, followed by underlying medical conditions like diabetes and hypertension. However, we compared its performance with individual methods and found that the ensemble methods, Random Forest and Voting Classifier, are better, achieving higher precision and Recall than individual ones. It demonstrates how machine learning can be used to identify high-risk patients or to figure out where resources should go.

4 Results

Based on key insights from analysis, it is found that Age and comorbidities are good predictors of survival of COVID-19. Feature importance plots, created from ensemble methods, such as Random Forest, indicated that Age was always the most significant predictor, followed by underlying medical

conditions like diabetes and hypertension. However, we compared its performance with individual methods and found that the ensemble methods, Random Forest and Voting Classifier, are better, achieving higher precision and Recall than individual ones. It demonstrates how machine learning can be used to identify high-risk patients or to figure out where resources should go.

5 Results

In this section, confusion matrices of various machine learning classifiers, such as Decision Tree, Random Forest, Logistic Regression, and Gradient Boost, are computed, and key performance metrics like true positives, false positives, false negatives, and true negatives are examined. The results display the strengths and weaknesses of each model to predict the COVID-19 mortality risk depending on the given dataset. The results highlight which models perform better in detecting high-risk patients to achieve that balance between accuracy and producing fewer errors.

5.1 Performance Overview

This insight is crucial, which is why the feature importance analysis of this model offers insight into what affects its prediction of COVID-19 mortality risk. It can be seen from Figure X that Patient Type was the most prominent predictor, revealing an evident influence on mortality risk. This corresponded with the clinical understanding that hospital or critical care patients are at higher risk. Pneumonia and Age were also high in the rankings, reinforcing well-known links to severe outcomes in COVID-19 cases. Elderly patients with pneumonia are much more vulnerable to weakened immune responses, for instance. More importantly, some essential features such as Hypertension and Diabetes are well-known comorbidities that exert a negative impact on the COVID-19 prognosis. Despite lower importance, features like Obesity and Chronic Conditions were also present in the model, improving its predictive effect. Surprisingly, there weren't too many features (excluding specific medical unit identifiers) that exhibited many contributions. Instead, patient-specific factors like Age and comorbidities significantly contributed to predicting outcomes. In addition to validating the clinical relevance of these factors, this analysis furnishes actionable insights for healthcare professionals to determine where to focus their efforts, focusing on high-risk patients where they can effectively influence weight reduction. Figure 3

indicates the feature importance bar graph used in this study.

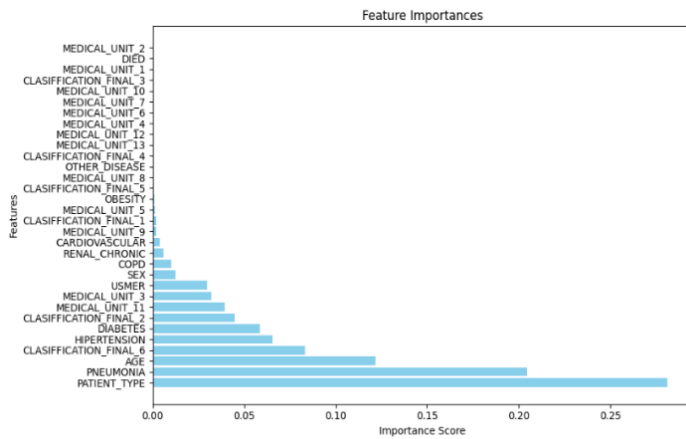


Figure 3. Feature Importance.

5.2 Model Performance Analysis

Confusion matrix results from all classifiers reveal the performance of all the classifiers in predicting probabilistic COVID-19 mortality risk. Figure 3 Comparison of K Meas for the SVM, Naïve Bayes, Decision Tree, and Random Forest Classifiers on the Diabetes Dataset shows how the Decision Tree Classifier, as shown in Figure 4 (a) and the Random Forest Classifier, as shown in Figure 4 (b) produce identical results classifying 13,161 true negatives (correctly identifying non-deceased patients) and 14,055 true positives (correctly identifying deceased patients). Both models, however, misclassified 1,786 cases as false positives (non-deceased patients predicted as deceased) and 966 cases as false negatives

(deceased patients predicted as non-deceased). The results suggest both models worked fairly well but could not get false positives to a minimum that would not waste resources.

The Logistic Regression Classifier, as shown in Figure 4 (c) performed best in terms of the highest true positive (14,183) and lowest false negative count (788). Thus, the model worked best at correctly identifying high-risk patients with a reasonable 'false alarm' count of 1,783. This balance makes the protocol very suitable for healthcare applications where the cost of missing high-risk patients is high. This model's ability to reduce false negatives is critical to making good, timely decisions.

Among other classifiers, the Gradient Boosting Classifier, as shown in Figure 4 (d) correctly identified lower-risk patients with the highest accurate negative count of 13,339 and the lowest false positive count of 1,608. Although it had 1,113 fewer false negatives, meaning not all high-risk patients were caught, it had fewer positives. In less critical applications, this tradeoff might be acceptable. Still, it could be problematic in healthcare, where it's often a matter of identifying which patients are at high risk.

The Alternative Logistic Regression Classifier, as shown in Figure 4 (e) attempted to minimize false positives and achieved the lowest false positives of 1,442, as shown in Figure 4 (f). But this meant we had 1,636 more false negatives and 13,335 fewer true positives. As a result, it is less appropriate as a healthcare application when the objective is to

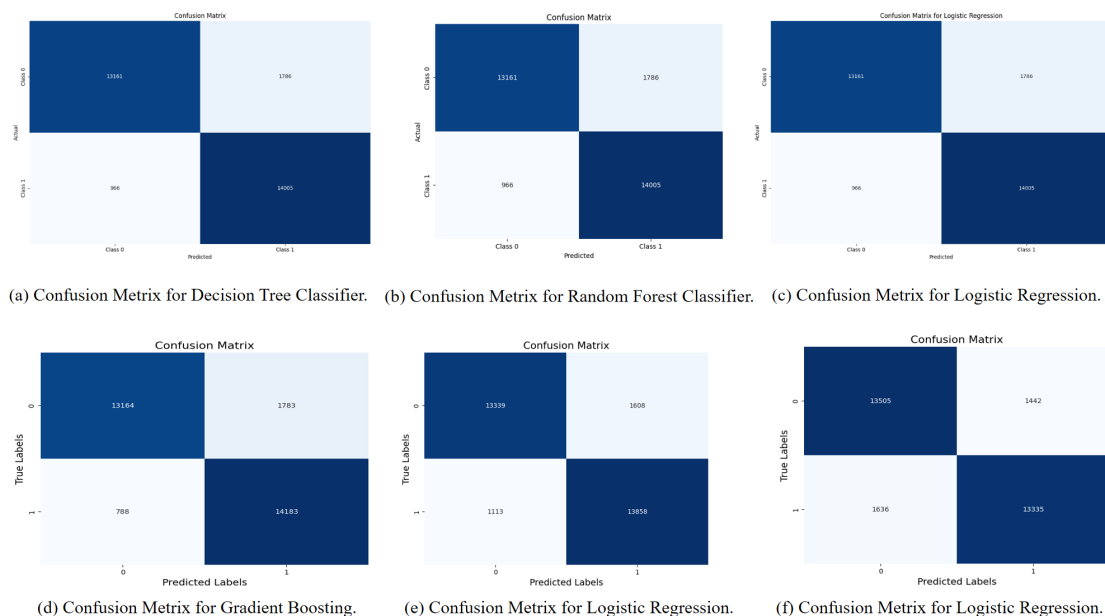


Figure 4. Confusion Matrix for all Models.

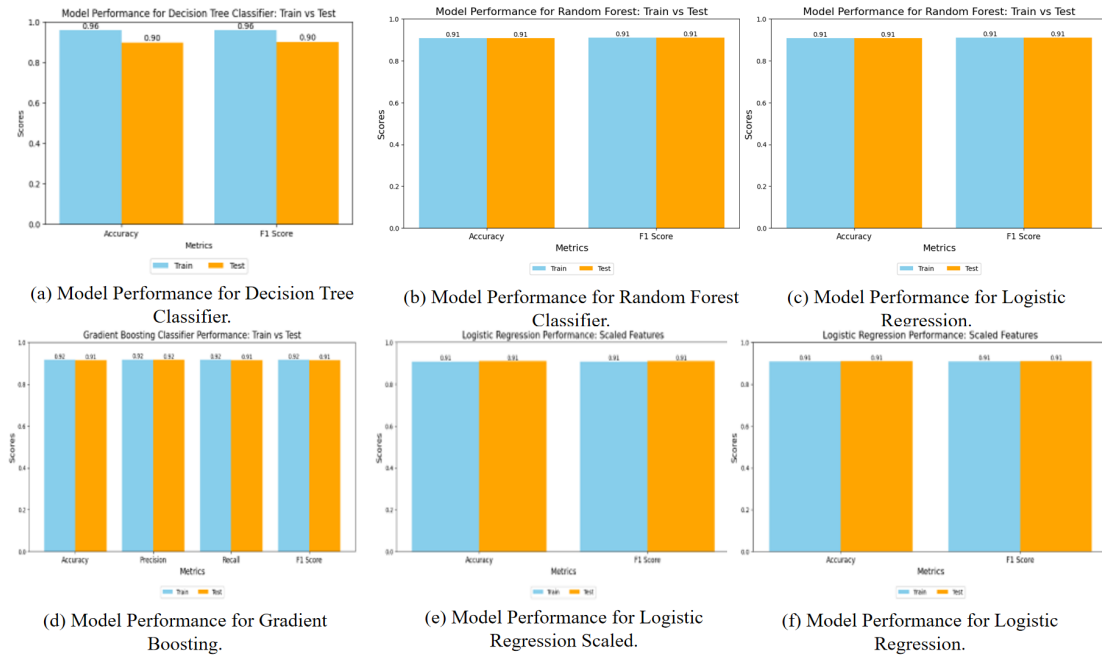


Figure 5. Model Performance.

recognize as many high-risk patients as possible.

Finally, considering both the ability to maximize true positives and minimize false negatives, the Logistic Regression Classifier, as shown in Figure 4 (c) was the best-performing model. It performs well enough to identify high-risk patients correctly, making it the best option for healthcare applications that use patient outcomes to life or death. Other models, e.g., Gradient Boosting, as shown in Figure 4 (d) and Decision Tree/Random Forest, as shown in Figure 4 (a) and (b), performed relatively well in some of these parts but less well in general more important aspects of healthcare scenarios. All model variations of confusion matrix is shown in Figure 4 (a) to (f).

Different machine learning classifiers were then evaluated for training and testing data sets based on their accuracy, F1 score, precision, and Recall. Figures 5 (a) to (f) visualize the strengths and limitations of each model as revealed by the results.

From the training given in the Decision Tree Classifier, as shown in Figure 5 (a), we observe good performance on the training dataset, with an accuracy and F1 score of 96%. On the test set, however, these metrics decreased slightly, with accuracy and F1 score dropping to 90%. This indicates a tendency to fit the model to the data slightly more than it should, with the model performing better on the training data than on the unseen data.

The Random Forest Classifier, as shown in Figure 5

(b) generally performed similarly for the training and Testing sets and achieved the same 91% accuracy and F1 score. This means that the model does not overfit the original data and can predict well on unseen data.

The Logistic Regression Classifier, as shown in Figure 5 (c) worked consistently and returned an accuracy and F1 score of 91% on both the training and testing datasets. The metrics are stable, which shows that logistic regression is a good model for this classification task in terms of stability and reliability.

Moving onto Figure 5 (d), the Gradient Boosting Classifier exhibited the highest accuracy and F1-score of 92% on the training dataset that stayed intact on the test dataset. In addition, this model showed excellent precision and Recall and was very useful for applications that require high sensitivity and specificity.

Finally, Figures 5(e) and 5(f) show that the Logistic Regression Classifier with Scaled Features returned the same accuracy and F1 score as the standard Logistic Regression model, with accuracy results of 91% on both training and testing datasets. Using feature scaling neither improved the model much nor kept it reliable and consistent.

Overall, the Gradient Boosting Classifier resulted in the best model performance with the highest accuracy, highest F1 score, highest precision, and highest Recall. The random forest classifier and logistic regression classifier performed very well, and they are both

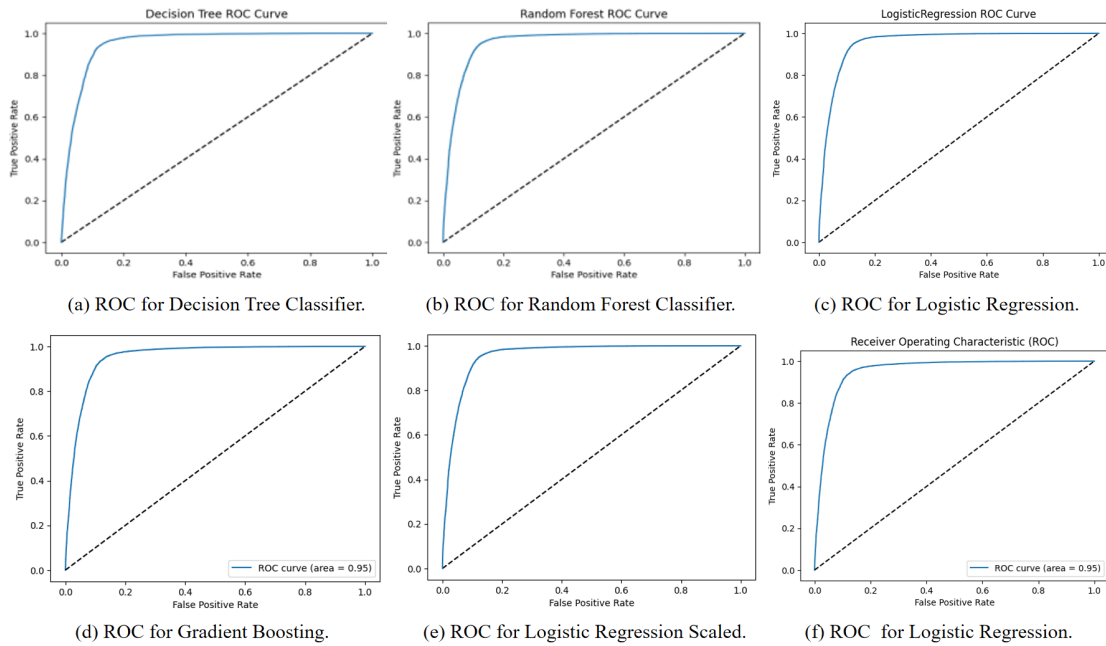


Figure 6. ROC Curve for all Models.

reliable choices. The Decision Tree Classifier also exhibited some overfitting, and feature scaling did not improve the performance of Logistic Regression. This study's findings indicate that the Gradient Boosting Classifier is the most successful model to predict a COVID-19 death risk. All models' model performance is shown in Figure 5(a-f).

The Receiver Operating Characteristic (ROC) curves of all classifiers show their discrimination ability to discriminate between positive (deceased) and negative (non-deceased) cases. Figures 6 (a) - (f) display the ROC curves of each classifier, which represent the tradeoff between the true positive rate (sensitivity) and the false positive rate. The classifiers are evaluated by the area under the curve (AUC) for the classifier.

Figure 6 (a) shows the Decision Tree Classifier, which has great discriminatory power (high AUC). Looking more closely at the earlier metrics, we can see that there is some slight overfitting, but overall, the curve indicates that the model does well at distinguishing high-risk from low-risk patients, assuming that generalizability is not substantially impacted by the overfitting.

The Random Forest Classifier displays an almost identical ROC curve, as shown in Figure 6 (b) with a similarly high AUC. This consistency alludes to the fact that Random Forest utilizes ensemble learning to decrease variance and provide reliable predictions.

The Logistic Regression Classifier, as shown in Figure 6 (c) also shows high AUC and is, therefore, a robust

classifier for discriminating between the two classes. Because logistic regression achieves a balanced tradeoff between sensitivity and specificity, it is a very good choice for the binary classification task in healthcare environments.

The Gradient Boosting Classifier, as shown in Figure 6 (d) had the highest AUC among all models. In addition, the ROC curve is closer to the top left corner than the others, indicating its superior ability to correctly recognize true positives while preventing too many false positives. As can be seen, Gradient Boosting is a very effective classifier for this dataset.

The ROC curves for all models are shown in Figure 6(a-f).

The logistic regression model with Scaled Features, as shown in Figure 6 (e) also gave a very similar AUC, indicating its consistent performance. The model was similar in its ability to discriminate classes whether or not feature scaling was employed.

Since the alternate Logistic Regression, as shown in Figure 6 (f) had the same ROC curve as the scaled Logistic Regression, scaling had no effect on performance. The model's stability and reliability stayed high, and the AUC remained the same.

Finally, conclusions are drawn based on the fact that all classifiers attained high AUC values, thus enabling them to make predictions. Of the committees, Gradient Boosting, as shown in Figure 6 (d) had the best performance, as shown in Figure 5, achieving the best

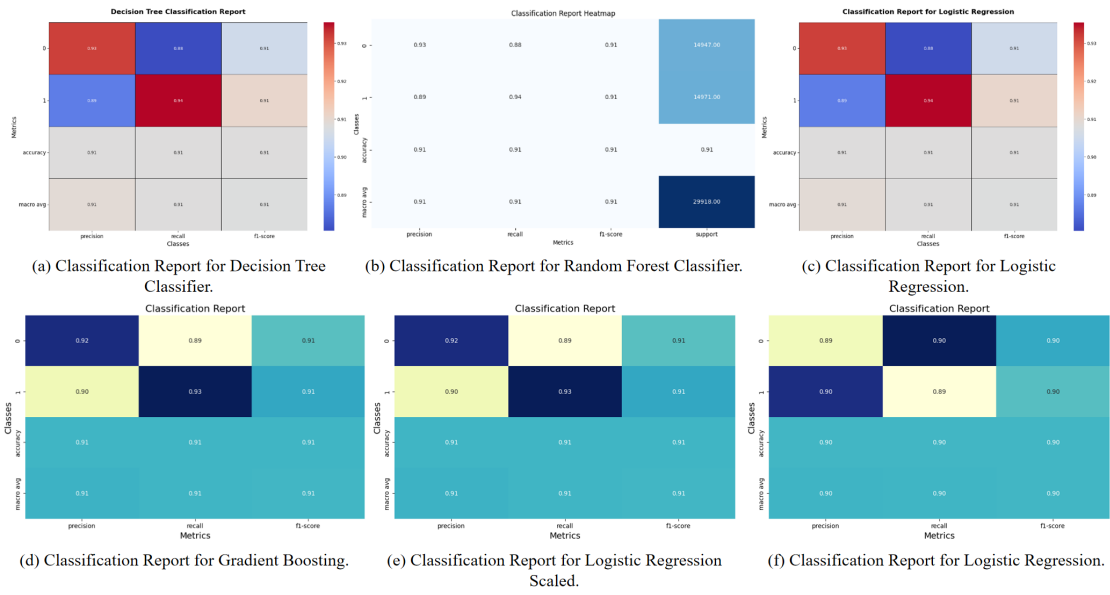


Figure 7. Classification Report for all Models.

compromise of sensitivity and specificity. On the other hand, logistic Regression models, as shown in Figure 6 (c), (e), (f) also performed very well and exhibited desirable properties for tasks that demand readable and consistent predictions. Gradient Boosting and Logistic Regression worked better, but not as robustly, as the Decision Tree and the Random Forest classifiers. In particular, these findings demonstrate the power of ensemble methods and logistic regression in predicting the risk of COVID-19 mortality.

The evaluated models' classification reports provide detailed insights into the performance metrics, including precision, Recall, and F1-score for classes (Class 0: Class 1: deceased) and Class 2: non-deceased. Moreover, the macro-averaged scores and the overall accuracy suggest the generality of the models.

The Accuracy of the Decision Tree Classifier, as shown in Figure 7 (a) was 91%, Class 0 had a Precision of 93%, Recall of 88%, and F1 score of 91%; while Class 1 had a Precision of 89%, Recall of 94%, and F1 score of 91%. These results show good performance but a slight decrease in Recall for Class 0, meaning some cases with zero were not recognized correctly.

Similarly, using the Random Forest Classifier, as shown in Figure 7 (b) resulted in an overall accuracy of 91%. Class 0 reported a precision of 93%, Recall of 88%, and F1-score of 91%, while Class 1 got a precision of 89%, Recall of 94%, and F1-score of 91%. Random Forest is consistent, splitting both classes equally and doing well on all metrics.

The Logistic Regression Classifier, as shown in Figure 7

(c) also produced an accuracy of 91%, with a precision of 93%, Recall of 88%, and F1-score of 91% for Class 0. Class 1 performance was 89% precision, 94% recall, and 91% F1 score. The results validate that Logistic Regression is a robust and consistent tool for predicting mortality from COVID-19.

Among all classifiers, the Gradient Boosting Classifier, as shown in Figure 7 (d) improves slightly in terms of accuracy (92 %), as it enables better balancing of precision, Recall, and F1-score on both classes. With Class 0, we got a precision of 92%, a recall of 89%, and an F1 score of 91%, and with Class 1, we got a precision of 90%, a recall of 93%, and an F1 score of 91%. Gradient Boosting is particularly effective for Class 1 because the improved Recall for Class 1 allows it to identify high-risk patients.

Similar observations were made from the Scaled Logistic Regression Classifier, which had the same overall accuracy as the unscaled Logistic Regression Classifier, as shown in Figures 7 (e) and (f). Precision (92%), Recall (89%), and F1-score (91%) of both Class 0 and Class 1 were balanced, while their f1 precision (93%) and f1 recall (94%) were even higher. This consistency showcases that the model's predictive performance was not significantly changed by the feature scaling.

The classification report for all the models is shown in Figure 7(a-f).

6 Discussion

To evaluate machine learning classifiers in identifying the risk of COVID-19 mortality, we analyzed their performance metrics, confusion matrices, ROC curves, and classification reports and observed each model's strengths and weaknesses. In the following, we discuss these findings and their implications for healthcare decision-making and resource allocation.

6.1 Model Strengths and Weaknesses

For both the training and testing phases, the Decision Tree Classifier achieved high values of accuracy and F1 scores (the latter equals 96% and 90%, respectively). However, the drop in performance is very slight between training and testing, which hints at some mild overfit in the training. Confusion matrixes indicated that the Decision Tree correctly diagnosed almost all of the deceased and non-deceased patients; however, 1,786 false positives misdiagnosed non-deceased patients as deceased, and 966 false negatives were diagnosed as non-deceased. The model, although strong, appears to have a higher false positive rate, implying that unnecessary healthcare resources will be allocated. The classification report corroborated this; the Recall serves 88% of non-deceased cases (Class 0), meaning we have a slight challenge in recognizing all low-risk patients.

Since this is an ensemble method, the Random Forest Classifier has slightly better stability than the Decision Tree. It returns almost similar results for both training and testing datasets (accuracy and F1 score of 91% for both). From the confusion matrix and the classification report, I got the same performance as with the Decision Tree: 1,786 false positives and 966 false negatives. However, compared to a single Decision Tree, Random Forest performs with reduced variance and reduces the risk of overfitting or generalizability.

Out of all models, the Logistic Regression Classifier showed a more balanced model in terms of accuracy and F1-s score of 91% for both training and testing datasets. Additionally, for health care domains where minimizing the number of high-risk cases missed is crucial, the confusion matrix counts for the lowest number of false negatives (788) and lower number of false positives (1,783), therefore, are very suitable. The performance in the classification report showed consistent precision, Recall, and F1-scores for both classes, which verify its robustness and reliability. Another advantage of Logistic Regression is that it's interpretable so that healthcare professionals may understand the impact of individual features on

mortality risk.

Moreover, the Gradient Boosting Classifier was found to be the best-performing model in terms of total accuracy (92%) and F1 scores for continuous and complete datasets. Across both classes, it had excellent precision, Recall, and a very low false positive rate (1,608). Based on the ROC curve, its outstanding performance was further illustrated by the curve approaching the top left corner well, implying good discriminatory power. This means Gradient Boosting is a great fit for healthcare applications where we are very sensitive (don't miss a high-risk patient) and somewhat sensitive (don't waste resources on unneeded tests). Nevertheless, it is computationally more complex than Logistic Regression, and thus, it wouldn't be used in a time-sensitive or resource-constrained environment for practical purposes.

For instance, for the Logistic Regression Classifier with Scaled Features, the accuracy and F1-scores were identical to Logistic Regression with Unscaled Features, both at 91%, respectively. Thus, feature scaling didn't significantly affect the model's performance, implying that Logistic Regression is intrinsically robust to this dataset's variances in feature magnitude. Though consistent, this is not improving much with scaling, and that might mean that simpler preprocessing steps would be enough for this model.

6.2 Comparison and Implications

Comparing all classifiers, we find a tradeoff between model complexity, interpretability, and performance. We find that Gradient Boosting is the best-performing model with the highest accuracy and precision but at the highest computational requirements. Similar to Logistic Regression, which strikes a good balance between strong performance and simplicity and interpretability, it is a good default setting for implementation in practice in healthcare settings. Between models like Decision Trees and the more computationally intensive Logistic Regression, Ensemble methods like Random Forests demonstrate how we can achieve stability and generalization with minimal effort.

All classifiers have exhibited very high AUC values, demonstrating that they hold great predictive power and could benefit healthcare by assisting decision-making. Despite a slight overfitting in the Decision Tree, there was reasonable consistency of

performance on the Logistic Regression across folds, validating that simpler models are better suited for use on real-world problems where generalizability cannot be compromised.

6.3 Real-World Application

Choosing a predictive model in healthcare settings, where decisions sometimes carry life and death consequences, must balance accuracy, interpretability, and computational efficiency. Being capable of minimizing false negatives, Gradient Boosting excellently handled the task of prioritizing high-risk patients so that they would be tended to as early as possible and the resources would be effectively allocated. This, however, can be computationally expensive, limiting its utilization in real-time settings or resource-constrained environments. Low false negative rate and interpretability of Logistic Regression as an alternative, offering bold decisions from healthcare providers based on the output from the model.

They also demonstrate the need to carefully select evaluation metrics. Although accuracy is a general performance metric, other metrics such as precision, Recall, and F1 score are more applicable to healthcare since the cost of false positives and false negatives is quite different. For example, false negatives, i.e., missing high-risk patients, can have dire consequences, and thus, Recall is a key metric when deciding if a model will fit the bill.

6.4 Future Directions

Our work demonstrates that machine learning can be employed for COVID-19 mortality risk prediction and several hypothesis untested regions. Incorporating other characteristics, like socioeconomic considerations or healthcare access, increases model performance and offers a better contextual representation of mortality risk. Constructing hybrid models that amalgamate the best of Gradient Boosting and Logistic Regression for constructing models will enhance predictive accuracy and help interpretability. The generalizability and practical applicability of the models evaluated in this thesis will be confirmed by further considering the models on other datasets and in real-world clinical environments.

7 Conclusion

This paper uses machine learning classifiers to predict COVID-19 mortality risk with potential applications in healthcare decision-making. The study performs

performance evaluation given models like Decision Tree, Random Forest, Logistic Regression, and Gradient Boosting, based on Accuracy, Precision, Recall, and F1 score with the support of confusion matrices and ROC curves. Gradient Boosting yielded the best accuracy (92%) and F1 scores and an outstanding balance between sensitivity and specificity by performing quantitative exercises most among all classifiers. Because of the low false positives and low false negatives in its ability to identify high-risk patients in such a manner, the system will be able to provide necessary, timely intervention in finding high-risk patients. However, its computational complexity may prevent applicability in real time, particularly in resource-constrained environments. Although slightly less accurate, Logistic Regression was consistently strong and interpretable, proving 91% correct and worst false negative. It is very simple and robust, making it very suitable for real-world applications where efficiency and transparency are of the essence. For instance, slightly more reliable models with balanced performance were ensemble methods (e.g., Random Forest). However, they did not provide significantly better results than simpler baseline models (e.g., Logistic Regression). Concurrently, the Decision Tree Classifier showed a small amount of overfitting, which signifies a better need for regularization techniques. This work emphasizes the need for a compromise between the performance, interpretability, and computational feasibility of the developed models in healthcare applications. The results highlight the potential use of machine learning to create the best possible resource allocation and better patient outcomes during pandemics. This work should be expanded to hybrid models, feature integration, and real-world clinical validation for predictive models in healthcare.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

- [1] Oran, D. P., & Topol, E. J. (2020). Prevalence of asymptomatic SARS-CoV-2 infection: A narrative review. *Annals of Internal Medicine*, 173(5), 362-367. [CrossRef]
- [2] Dabbah, M. A., Reed, A. B., Booth, A. T., Yassaee, A., Despotovic, A., Klasmer, B., ... & Mohan, D. (2021). Machine learning approach to dynamic risk modeling of mortality in COVID-19: a UK Biobank study. *Scientific reports*, 11(1), 16936. [CrossRef]
- [3] Oran, D. P., & Topol, E. J. (2021). Prevalence of asymptomatic SARS-CoV-2 infection. *Annals of Internal Medicine*, 174(2), 286-287. [CrossRef]
- [4] Byambasuren, O., Cardona, M., Bell, K., Clark, J., McLaws, M. L., & Glasziou, P. (2020). Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: systematic review and meta-analysis. *Official Journal of the Association of Medical Microbiology and Infectious Disease Canada*, 5(4), 223-234. [CrossRef]
- [5] Cao, Y., Hiyoshi, A., & Montgomery, S. (2020). COVID-19 case-fatality rate and demographic and socioeconomic influencers: Worldwide spatial regression analysis based on country-level data. *BMJ Open*, 10(11), e043560. [CrossRef]
- [6] Atkins, J. L., Masoli, J. A., Delgado, J., Pilling, L. C., Kuo, C. L., Kuchel, G. A., & Melzer, D. (2020). Preexisting comorbidities predicting COVID-19 and mortality in the UK biobank community cohort. *The Journals of Gerontology: Series A*, 75(11), 2224-2230. [CrossRef]
- [7] Li, B. (2020). The association between symptom onset and length of hospital stay in 2019 novel coronavirus pneumonia cases without epidemiological trace. *Journal of the National Medical Association*, 112(5), 516-517. [CrossRef]
- [8] Rechtman, E., Curtin, P., Navarro, E., Nirenberg, S., & Horton, M. K. (2020). Vital signs assessed in initial clinical encounters predict COVID-19 mortality in an NYC hospital system. *Scientific reports*, 10(1), 21545. [CrossRef]
- [9] Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., ... & Cao, B. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The lancet*, 395(10229), 1054-1062. [CrossRef]
- [10] Foy, B. H., Carlson, J. C., Reinertsen, E., Valls, R. P. I., Lopez, R. P., Palanques-Tost, E., ... & Higgins, J. M. (2020). Association of red blood cell distribution width with mortality risk in hospitalized adults with SARS-CoV-2 infection. *JAMA network open*, 3(9), e2022058-e2022058. [CrossRef]
- [11] Clift, A. K., Coupland, C. A., Keogh, R. H., Diaz-Ordaz, K., Williamson, E., Harrison, E. M., ... & Hippisley-Cox, J. (2020). Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study. *bmj*, 371. [CrossRef]
- [12] Knight, S. R., Ho, A., Pius, R., Buchan, I., Carson, G., Drake, T. M., ... & Harrison, E. M. (2020). Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. *bmj*, 370. [CrossRef]
- [13] Williamson, E. J., Walker, A. J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C. E., ... & Goldacre, B. (2020). Factors associated with COVID-19-related death using OpenSAFELY. *Nature*, 584(7821), 430-436. [CrossRef]
- [14] Yan, L., Zhang, H. T., Goncalves, J., Xiao, Y., Wang, M., Guo, Y., ... & Yuan, Y. (2020). An interpretable mortality prediction model for COVID-19 patients. *Nature machine intelligence*, 2(5), 283-288. [CrossRef]
- [15] Hamer, M., Kivimäki, M., Gale, C. R., & Batty, G. D. (2020). Lifestyle risk factors, inflammatory mechanisms, and COVID-19 hospitalization: A community-based cohort study of 387,109 adults in UK. *Brain, behavior, and immunity*, 87, 184-187. [CrossRef]
- [16] Redondo, E., Rivero-Calle, I., Mascarós, E., Ocaña, D., Jimeno, I., Gil, Á., ... & Martínón-Torres, F. (2024). Respiratory syncytial virus vaccination recommendations for adults aged 60 years and older: the pneumoexperts prevention group position paper. *Archivos de Bronconeumología*, 60(3), 161-170. [CrossRef]
- [17] Estes, C., Razavi, H., Loomba, R., Younossi, Z., & Sanyal, A. J. (2018). Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. *Hepatology*, 67(1), 123-133. [CrossRef]
- [18] Adadi, A., Lahmer, M., & Nasiri, S. (2022). Artificial Intelligence and COVID-19: A Systematic umbrella review and roads ahead. *Journal of King Saud University-Computer and Information Sciences*, 34(8), 5898-5920. [CrossRef]
- [19] Jose, R., Syed, F., Thomas, A., & Toma, M. (2024). Cardiovascular health management in diabetic patients with machine-learning-driven predictions and interventions. *Applied Sciences*, 14(5), 2132. [CrossRef]
- [20] Rehman, A., Noor, F., Janjua, J. I., Ihsan, A., Saeed, A. Q., & Abbas, T. (2024). Classification of Lung Diseases Using Machine Learning Technique. 2024 *International Conference on Decision Aid Sciences and Applications*, 1-7. [CrossRef]
- [21] Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, 14(1), 6086. [CrossRef]



Muhammad Adnan Hasnain is currently pursuing his graduate studies in Information Engineering, Computer Science, and Mathematics at the University of L'Aquila, Italy. His academic interests lie in the fields of computer science and engineering, with a focus on innovative technological solutions and advanced computational techniques.

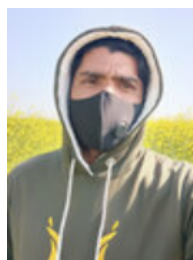
(E-mail: muhammadadnan.hasnain@graduate.univaq.it)



Tahir Abbas holds a Ph.D. in Computer Science from the National College of Business Administration & Economics in 2020. His research interests encompass Artificial Intelligence (AI), Bioinformatics, Computer Vision, Machine Learning (ML/DL), Federated Learning, and Explainable AI. Dr. Abbas has extensive experience in data science, research development, and scientific writing. (E-mail: drtahirabbas@t.edu.pk)



Jamshaid Iqbal Janjua, Ph.D. (Senior Member, IEEE) is a Data Scientist and Researcher with expertise in Machine Learning, Artificial Intelligence, and Data Science. He holds a Ph.D. in Computer Science, specializing in Machine Learning and Artificial Intelligence. Dr. Janjua also holds a Master of Science in Computer Science (MSCS) from the University of Central Punjab (2009) and a Master in Business Administration (MBA) in Human Resource Management from the Institute of Management Sciences (Pak-Aims) (2007). He is currently working as the Senior Manager of Research at KICS. (E-mail: jamshaid.janjua@kics.edu.pk)



Sunawar Khan is currently pursuing a Master's degree in Computer Science (MSCS) at the International Islamic University, Islamabad, Pakistan. His research interests focus on Artificial Intelligence (AI), Machine Learning (ML), and Smart Grids, with a particular emphasis on solving challenges in cybersecurity, IoT, and energy management. (E-mail: sunawarkhan@gmail.com)