Final Presentation
on
Development of Artificial Intelligence Based Computation Model for Prediction of COVID-19 Patient Outcome

Presented by: **Sayani Das**

(Reg.No.2021BM12)

Under the Supervision of: **Dr. Abhishek Kumar Tiwari**

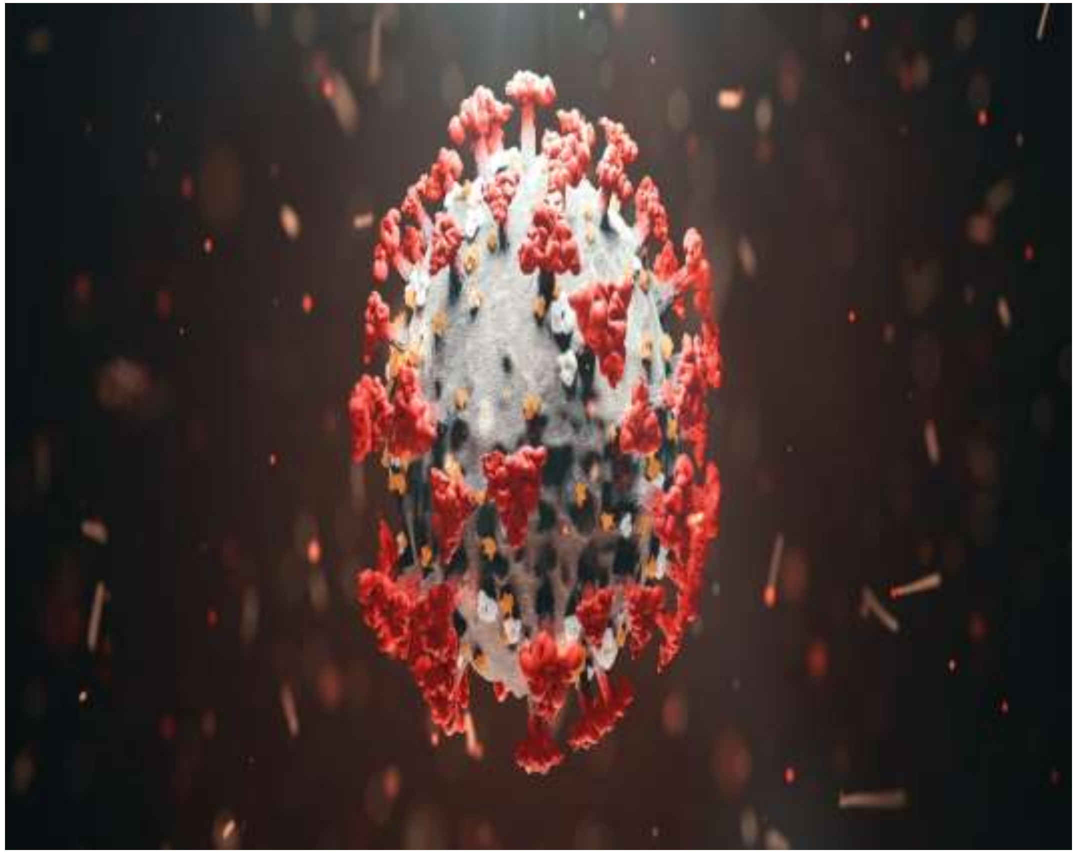
Applied Mechanics Department
Motilal Nehru National Institute of Technology Allahabad,
Prayagraj, Uttar Pradesh-211004

OUTLINE

- **Introduction**
- **Motivation**
- **Literature survey**
- **Research Gap**
- **Objective**
- **Methodology**
- **Results**
- **Conclusion**
- **Future Scope**
- **References**

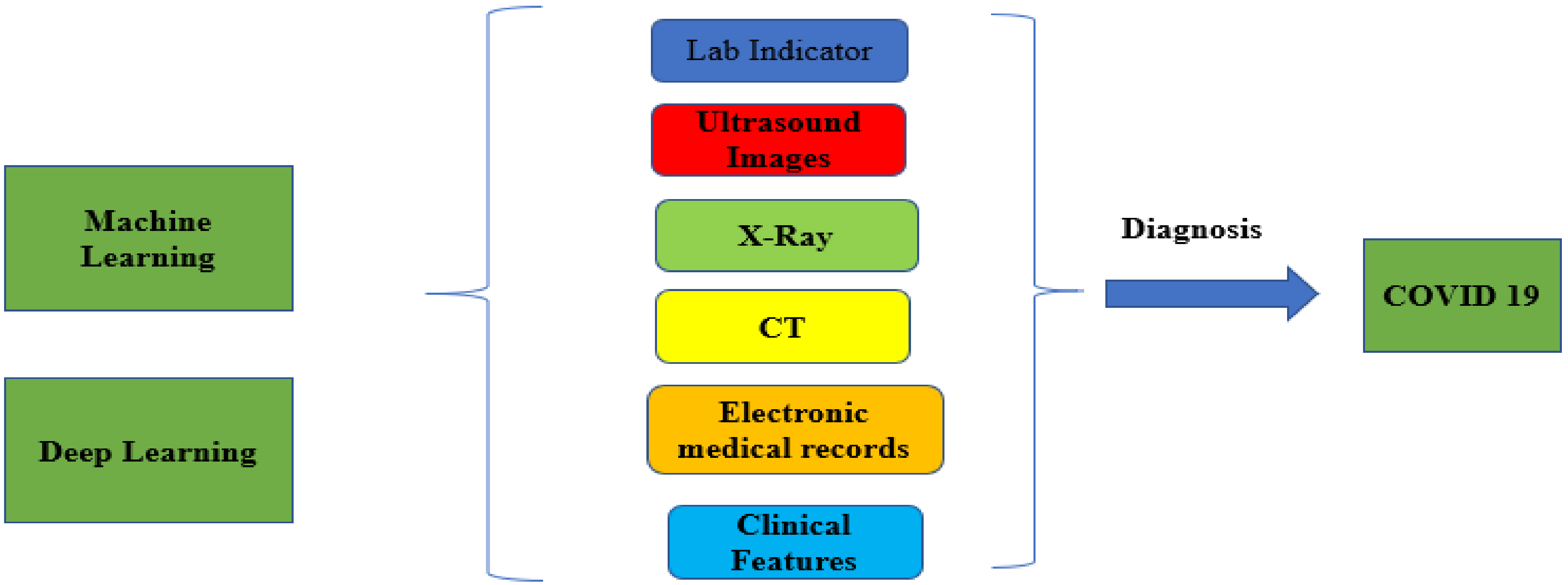
INTRODUCTION

- COVID-19, short for Coronavirus Disease 2019, is an infectious disease caused by the novel coronavirus SARS-CoV-2.
- It first started in December 2019 in the Chinese city of Wuhan, in the province of Hubei, and has since spread worldwide.
- On January 30, 2020, the World Health Organisation (WHO) declared it a Public Health Emergency of Worldwide Concern. On March 11, 2020, the WHO classified it as a pandemic.



Contd…

- Predicting COVID-19 patient outcomes using machine learning is a significant area of research and application.
- Machine learning techniques can help healthcare professionals identify and understand factors that contribute to different outcomes for COVID-19 patients, such as hospitalization, disease progression, and mortality.



Artificial intelligence for COVID-19 diagnosis

Contd…

The COVID-19 virus, SARS-CoV-2, has undergone a number of variations as of my knowledge's in September 2021. Prior until that time, the following significant varieties have been discovered:

WHO label	Lineage	Earliest documented samples
Alpha	B.1.1.7	United Kingdom, Sep-2020
Beta	B.1.351	South Africa, May-2020
Gamma	P.1	Brazil, Nov-2020
Delta	B.1.617.2	India, Oct-2020

Contd…

USES OF COVID-19 BIOMARKERS

- Early diagnosis of disease.
- Confirmation and Classification of disease severity.
- Framing hospital admission requirements.
- Identification of high risk cohort.
- Predicting Outcome

Covid-19 Biomarkers
IL-6 D-Dimer Platelet- Count ESR HS-CRP
Basophils
Neutrophils
Lymphocytes
Monocytes

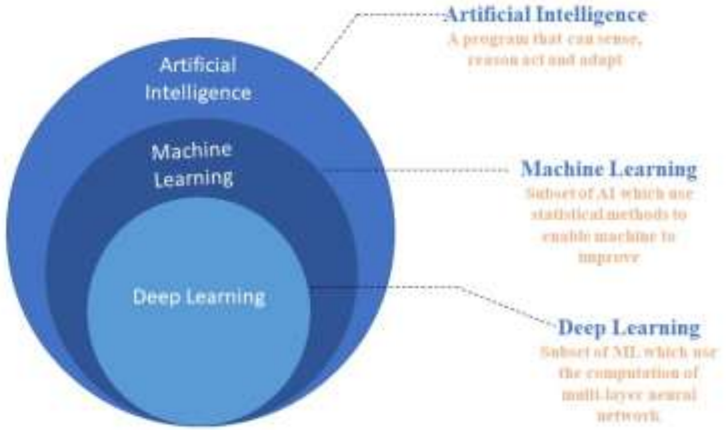
Contd…

ARTIFICIAL INTELLIGENCE (AI)

AI is basically that field of computer science which emphasizes on the creation of intelligent machines which can work and react like humans.

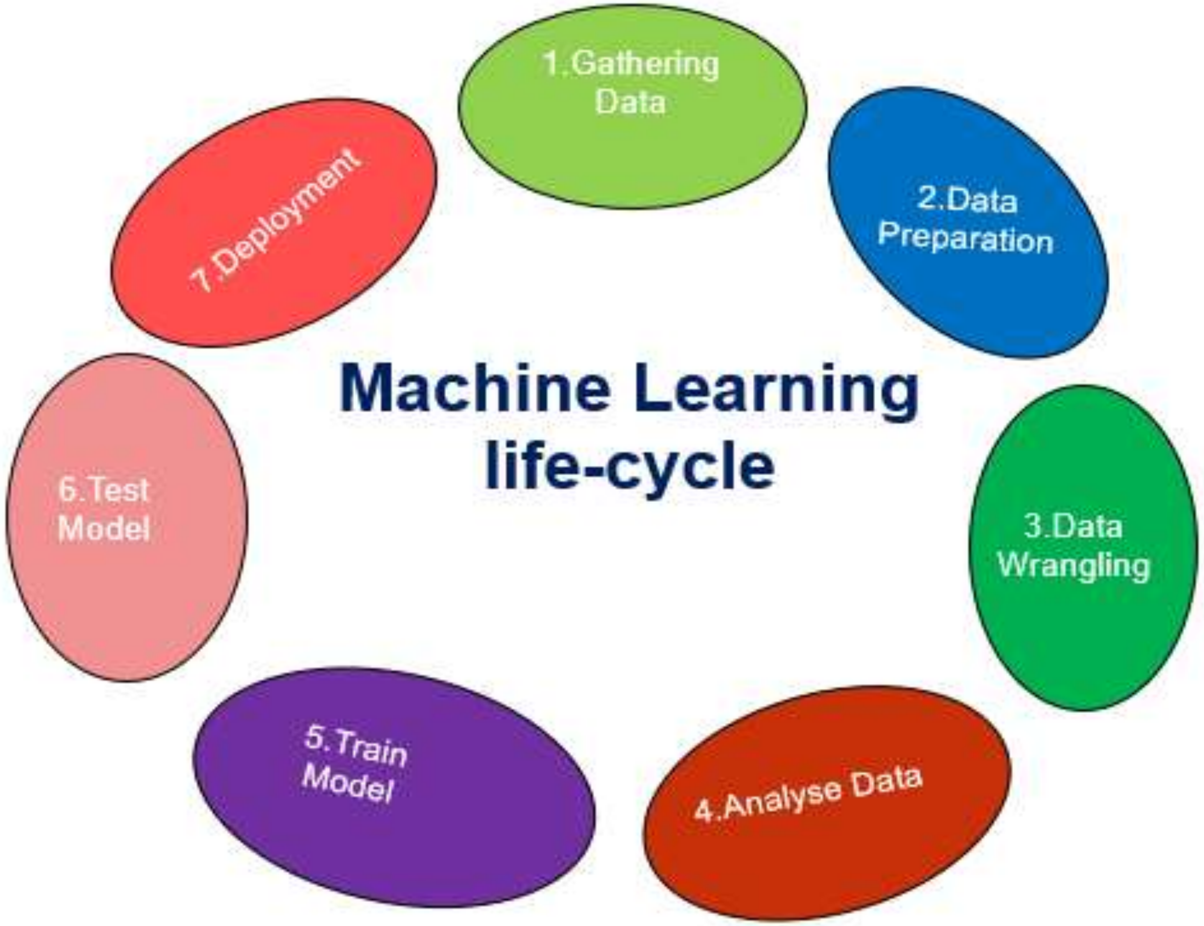
WHY AI?

- Widely used in banking and finance industry.
- Important feature of medical science (virtual personal he
- Perfect for heavy industry.
- A great help for humans, and many more.
- Deep Leaning (DL) and Machine Learning (ML) are the
- Deep Learning comes when Machine Learning fails.



Contd…

7 STEPS IN MACHINE LEARNING



/.

Contd…

MOTIVATION

The main Contributions of this study will be:

- Predicting the outcome of COVID-19 patients can aid healthcare professionals in making informed decisions regarding patient management and resource allocation.
- Accurate prediction of patient outcomes can support the development of personalized treatment strategies.
- Add information to the main points of current research to create the basis for ongoing research.
- By identifying specific biomarkers associated with survival or non-survival, can be designed to improve patient outcomes and increase the chances of survival.

LITERATURE REVIEW

Authors (Yr.)	Methodology	Main Findings	Outcome
Zhendong Xiao 2021[1]	ML techniques: Logistic regression, Neural networks , Light GBM, and Decision trees.	• The most predictable of the five models was the light GBM. • Age group as well as being admitted to a critical-care unit all became more impact on the chance of mortality.	The study found that the highest age group (>80) had the highest risk of death due to COVID-19.

LITERATURE REVIEW

Authors (Yr.)	Methodology	Main Findings	Outcome
Deif et.al(2021)[2]	Sample:550 sample used. Clinical Features:38 clinal features were collected.	The study demonstrated that machine learning models such DT, RF, SVM, KNN, ANN, and Xgboost were used to predict the seriousness of patients for ICU allocation.	The Xgboost classifier showed extremely precise accuracy of 97% in predicting mild patients.

LITERATURE REVIEW

Authors (Yr.)	Methodology	Main Findings	Outcome
Statsenko et al.(2021)[4]	Biomarkers used: Ferritin, Fibrinogen, Lactate dehydrogenase, CRP. The biomarkers were checked both of patients(ICU and non-ICU).	In accordance with the study, ICU patients much higher levels of several biomarker including CRP, LDH, Ferritin, and Fibrinogen to the non-ICU patients.	• There is a possibility for loss, which may necessitate admission to an intensive care unit (ICU). • The (AUC) score was found to be 0.86 in the study.

LITERATURE REVIEW

Authors (Yr.)	Methodology	Main Findings	Outcome
Schöning et al.(2021)[7]	• Dataset used: 287 Patients dataset. • Biomarkers: D-dimer, LDH value, and neutrophil-to-lymphocyte	• Admission into critical care units. • The factors that were most significant that determined the risk score were features including gender, CRP, and hemoglobin levels.	The using of Random Forest model, the greatest AUC score was obtained, which was 0.98.

LITERATURE REVIEW

Authors (Yr.)	Methodology	Main Findings	Outcome
Rahman et al.(2023)[5]	It requires using datasets collected from 930 COVID-19 patients who were admitted in Italy.	Identification of Low or High Levels of Emergency Patients. Identified with Covid-19 and their treatment With a Low or High Risk for the Covid-19 Variation..	A machine learning module was developed with an accuracy: 89.03% Sensitivity: 90.44% F1-score:89.03%

RESEARCH GAP

- Prediction of COVID-19 patients outcome (Survival vs Not-Survival)
- Comprehensive data analyzing and learning models and also building models that can rapidly modify and update predictions(Survival and Non-Survival).
- Computational to directly dataset and predict outcome based on Day-wise Biomarkers.

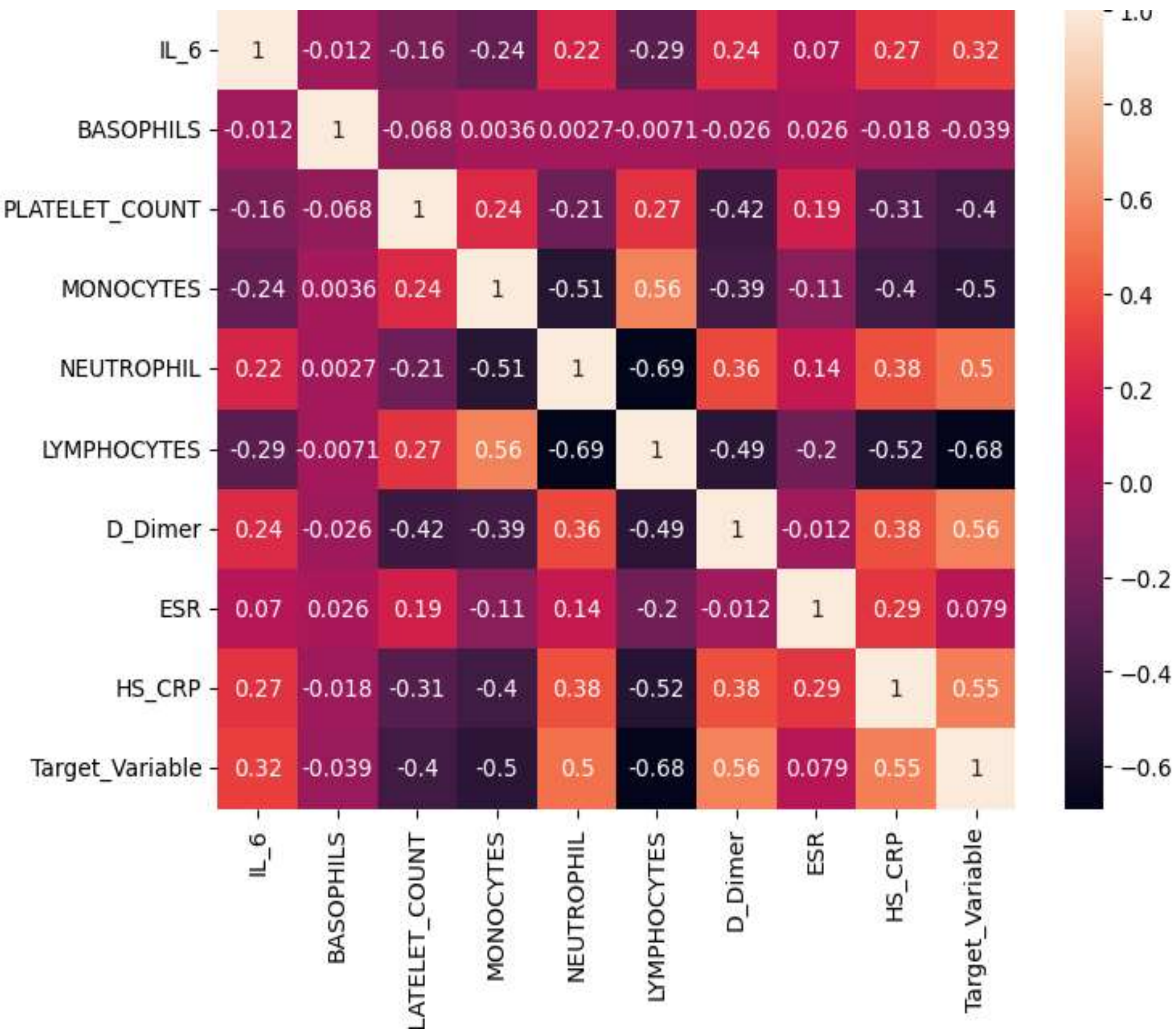
OBJECTIVES

- Prediction of the COVID-19 patients outcome.
- Collecting large amount of unprocessed Dataset.
- Selecting the most prominent biomarkers as per literature.
- Training and testing these biomarkers on 14 different models and calculating their accuracies.
- Based on the clinical parameters development and implementation of the computation model which predicts patients outcome i.e., Survival, and non-survival of the patients.

Methodology

Correlation Matrix

A correlation matrix is a statistical technique used to evaluate the relationship between two variables in a data set. The matrix is a table in which every cell contains a correlation coefficient, where 1 is considered a strong relationship between variables, 0 a neutral relationship and -1 a not strong relationship.



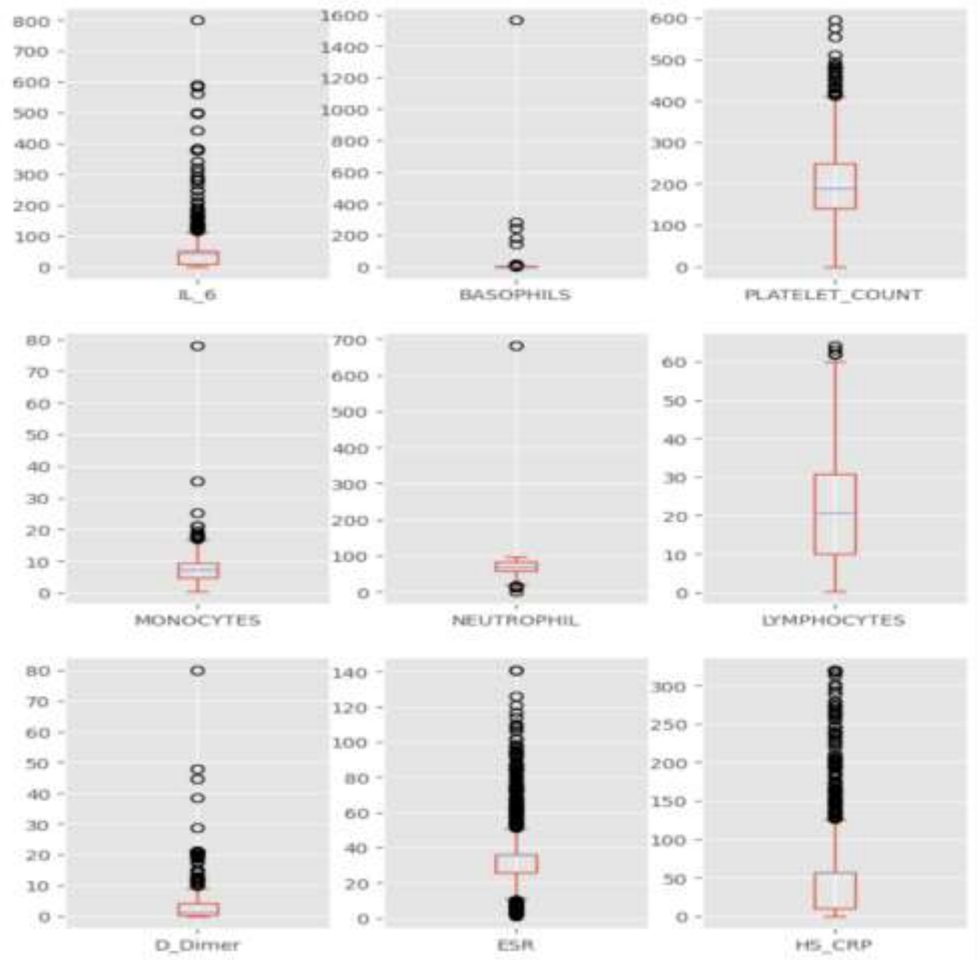
Handling Imbalanced Dataset

- After the dataset was cleaned, the sample size was reduced to 1578 cases. There were 789 and 219 samples in the survival and death classes, correspondingly. This dataset is imbalanced.

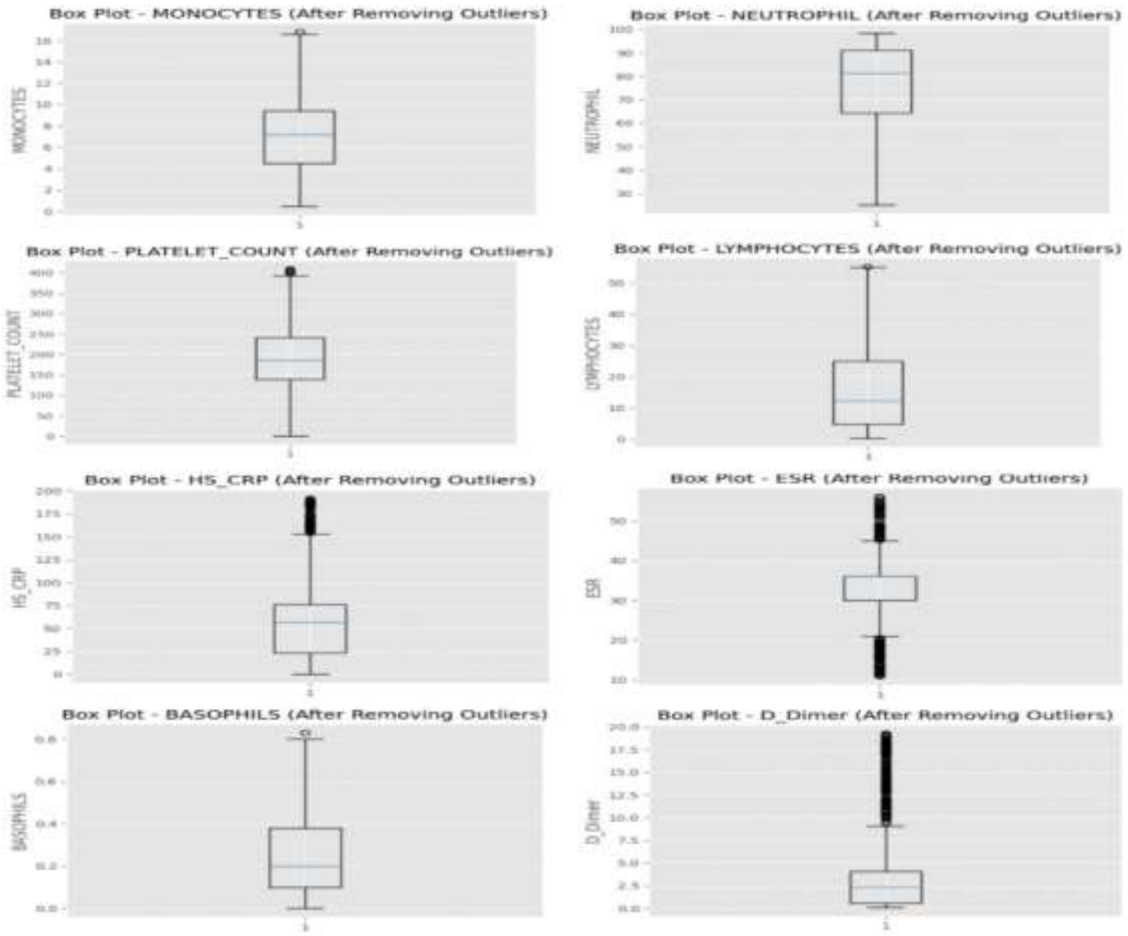
- To build the balanced dataset, Synthetic Minority Oversampling Technique (SMOTE) was implemented. To achieve balance, this method artificially generates samples from the minority class.

- A balanced dataset makes it possible for the classifier to make predictions about both classes.

Box-Plot to show the Outlier with Clinical Features(Before)



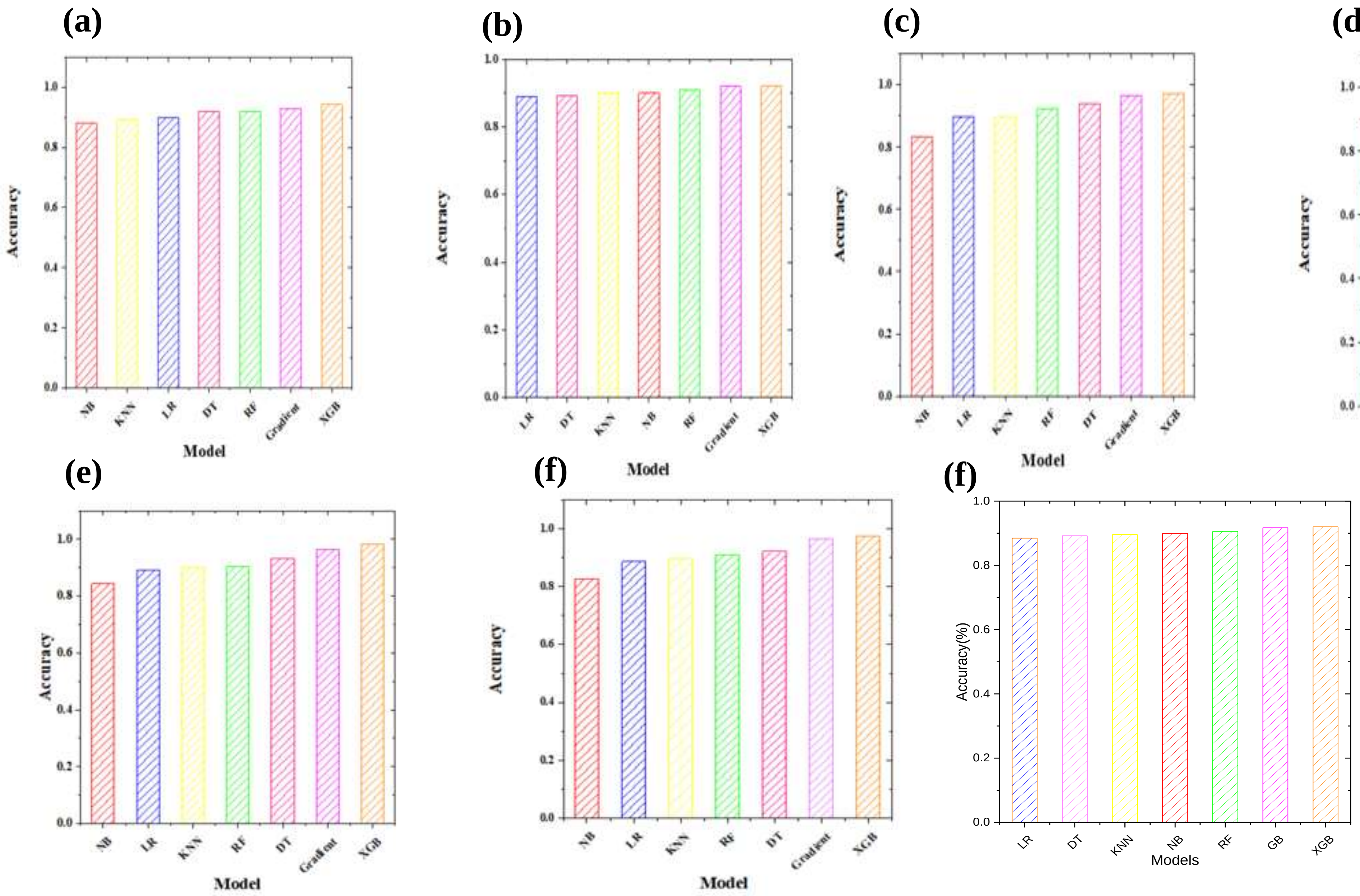
Box-Plot to show the Outlier with Clinical Features (After)



Results

- Our research showed that it is possible to predict COVID-19 patients' survival using their clinical characteristics.
- High accuracy, sensitivity, and specificity were achieved by the machine learning models, demonstrating their promise as prognostic tools.
- Additionally, the feature importance analysis demonstrated the major contributions of particular clinical indicators, underlining their function in the prediction of patient outcomes.

Accuracy Comparison of different Algorithms

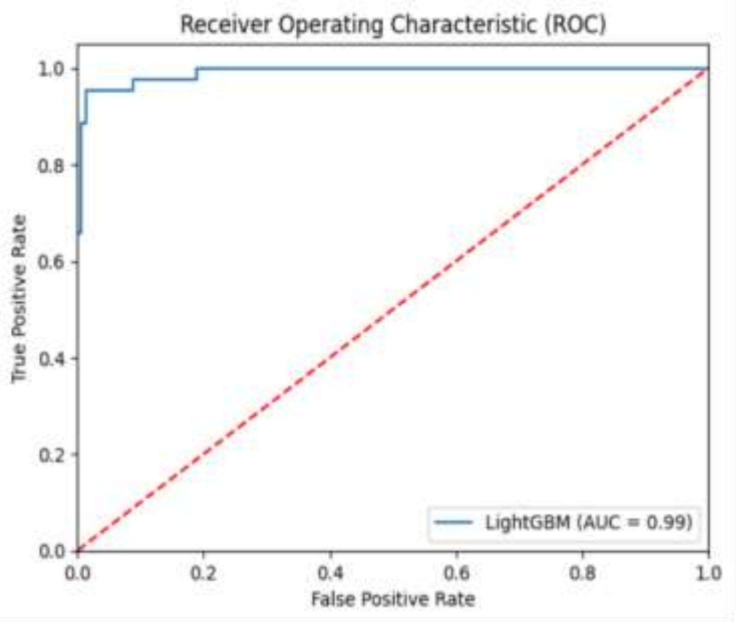


(a) 20% Training
(b) 30% Training
(c) 40 % Training
(d) 50% Training
(e) 60% Training
(f) 70% Training
(g) 80% Training

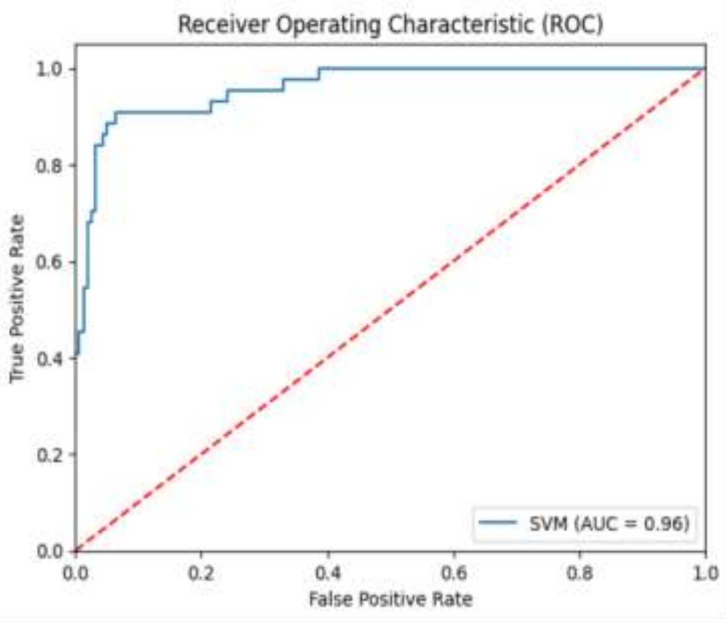
Performance of the different models

Model	Accuracy	Recall	Precision	F1 Score	AUC
Logistic Regression	87.12%	88.34%	96.17%	92.07%	90%
Random Forest	87.62%	77.77%	84.34%	80.33%	98%
Decision Tree	81.68%	75.53%	73.73%	74.33%	78%
Naïve Bayes	81.68%	83.46%	75.58%	77.51%	88%
K-Nearest Neighbor	85.15%	74.59%	79.72%	76.62%	86%
Xgboost	88.12%	79.67%	84.34%	81.63%	94%
Gradient Boost	87.13%	78.24%	82.68%	80%	94%
Adaboost	85.15%	77.76%	78.68%	78.20%	91%
Catboost	87.13%	78.24%	82.68%	80.10%	94%
Extratrees	87.62%	76.70%	84.34%	80.30%	92%
LSTM	93.7%	84.67%	72.56%	85.43%	91%
SVM	91.58%	68.18%	90.9%	77.92%	96%
1-D CNN	90.5%	84.09%	72.54%	77.89%	87.6%
LightGBM	98%	94.13%	95.45%	93.1%	99%

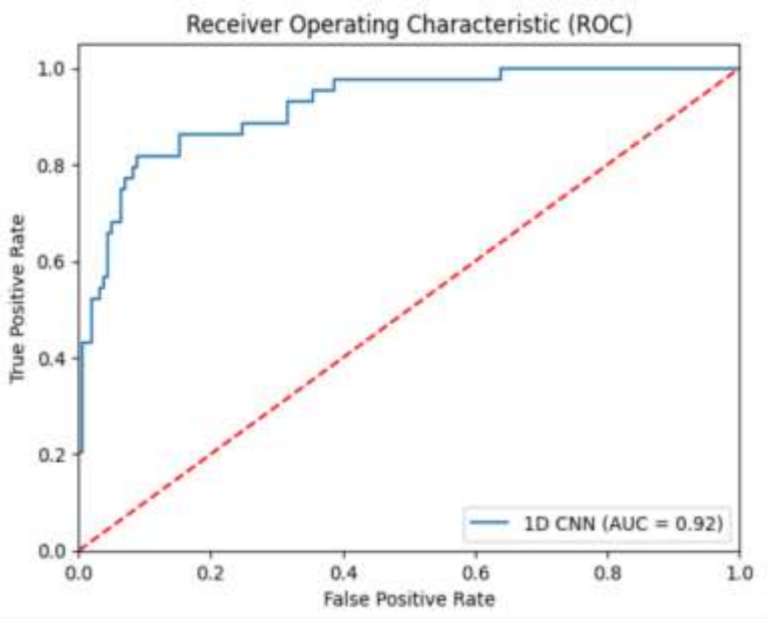
Evaluation AUC-ROC



(a) AUC-ROC curve for Light GBM



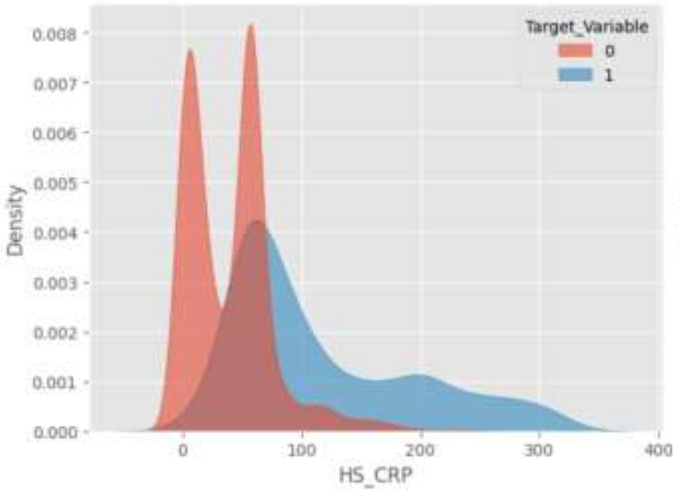
(b) AUC-ROC curve for SVM



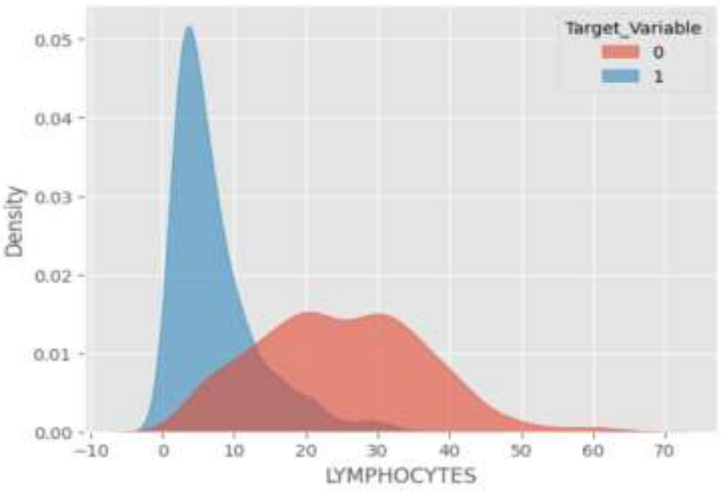
(c) AUC-ROC curve for 1D CNN

- A perfect classifier would have an AUC of 1, and it would always give a positive occurrence a higher score than a negative occurrence.
- In addition, a perfect classifier would not have any false positives or false negatives. Figure illustrates the area under the curve (AUC) as well as the receiver operating characteristic (ROC) curve for each of the machine learning algorithms that were applied during the process of developing our model.

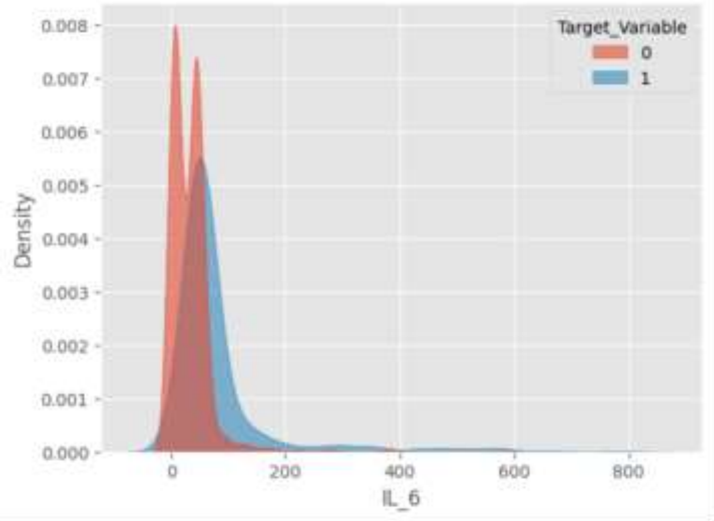
Feature and Density Plot



(a) Hs-CRP



(b) LYMPHOCYTES

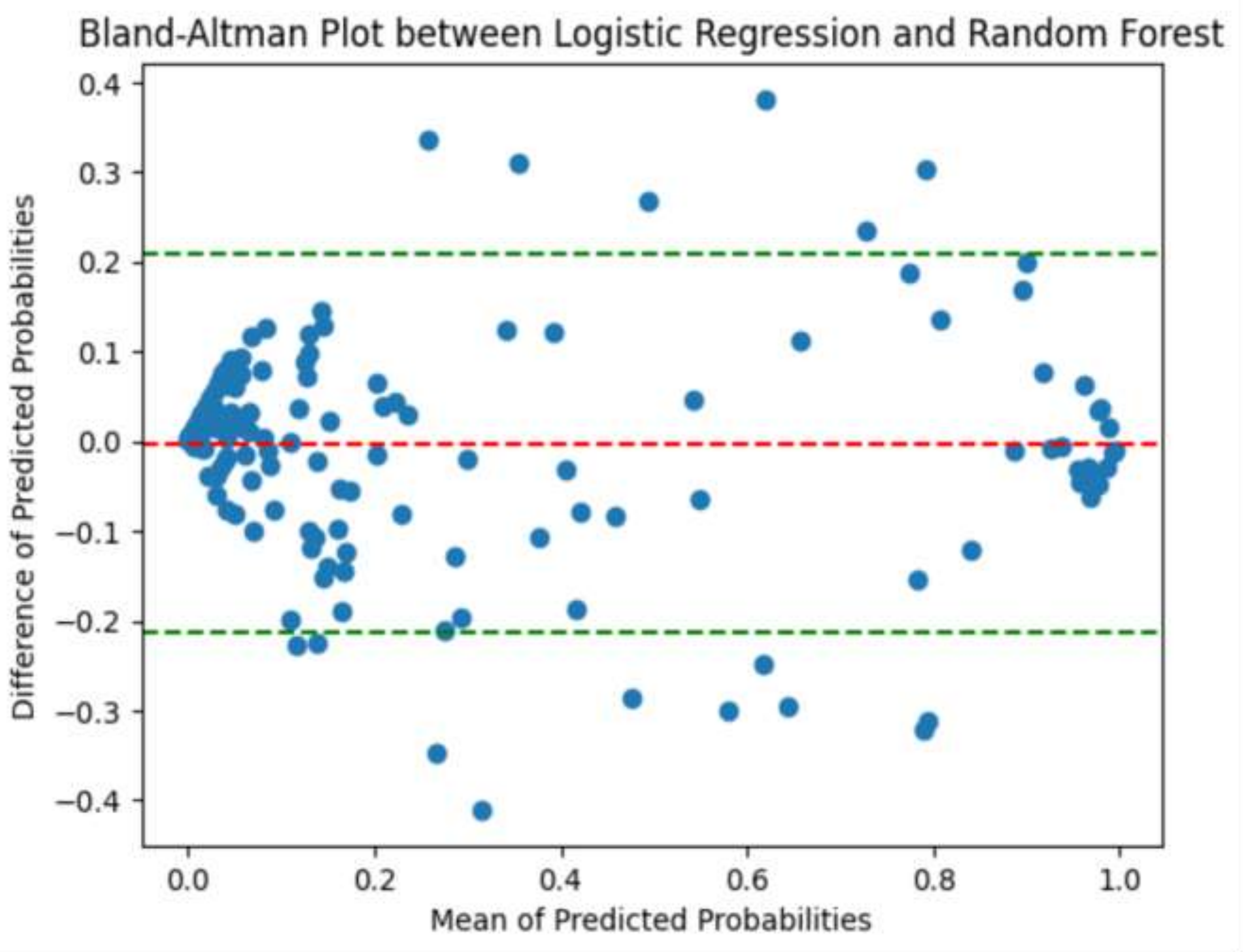


(c) IL-6

- We have the feature density histograms of surviving and death patients to assist us understand how various clinical factors differ in predicting patient death and survival.
- The value of HS_CRP was found to be significantly greater in patients who did not survive compared to individuals who survived.
- Additionally, it can be shown that the lymphocyte (%) value is higher for patients who survived in comparison to patients who died.

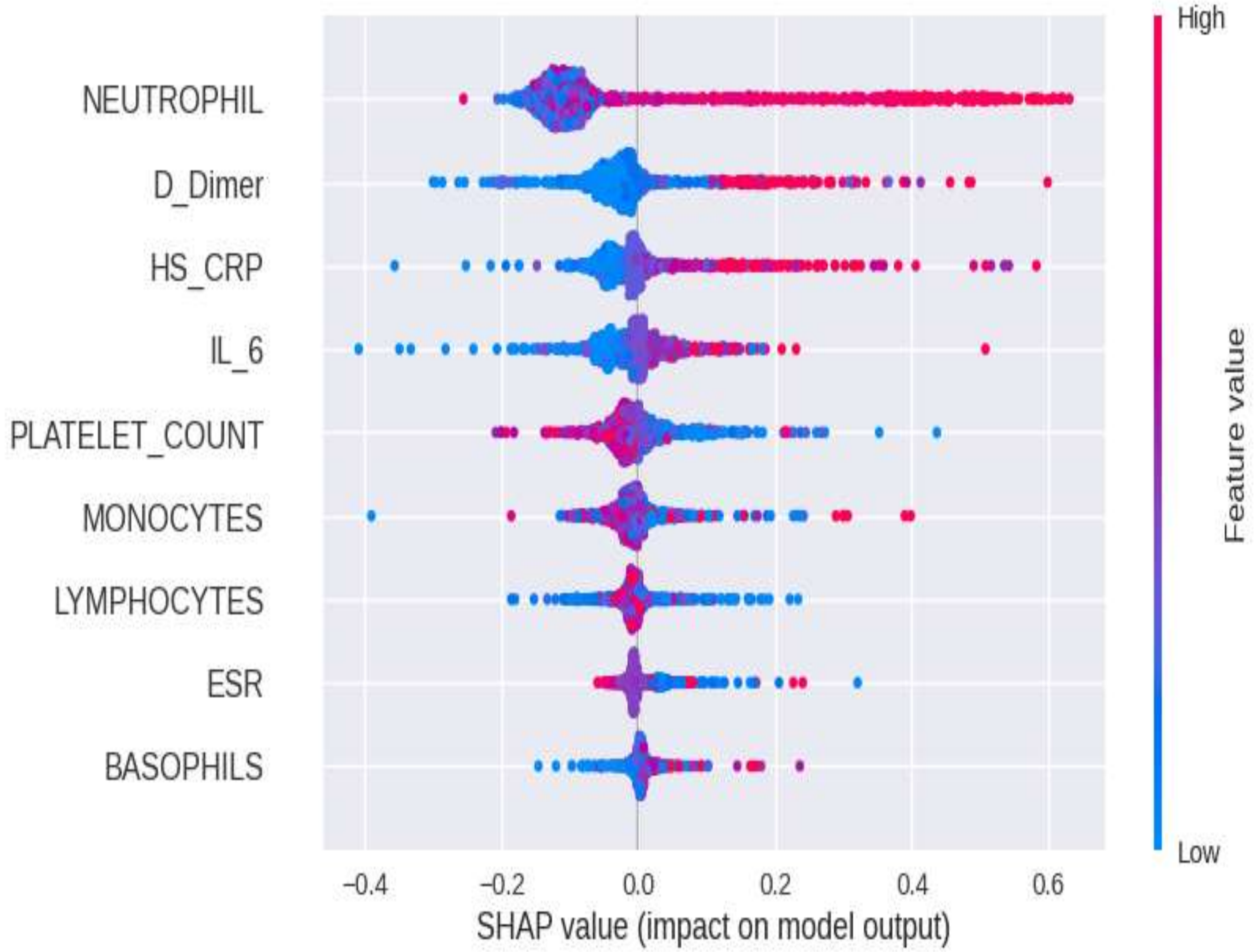
Bland Altman Analysis:

Analysis methods such as the Bland-Altman analysis are commonly used in the area of biomedical research. This method is used to examine the degree of agreement between two measurements that are taken on a continuous scale.



Shapley Additive Explanation(SHAP):

Game-theoretic classifier explanation tool SHAP is prominent. The contribution of each feature to a prediction determines its Shapley value. SHAP Beeswarm plot for Random Forest classifier. The beeswarm graphic provides a rich global interpretation summary for the random forest classifier. Each plot dot represents a data point. Feature names are ordered by importance on the y-axis. SHAP values are on the x-axis. Feature values are shown by colour gradient. The graphic contains more feature data for the redder spots.



CONCLUSION

- In this present study, we created a prediction algorithm with the intention of assisting the healthcare system in improving clinic decision-making and concentrating on COVID-19-infected patients.
- Patients who have been diagnosed with COVID-19 may reap benefits from the exploitation of the identified clinical indicators as valuable biomarkers for the purpose of risk classification and for monitoring the progression of disease.
- Before the models can be employed in clinical settings in the real world, they need to be developed and enhanced; however, this can only be done if additional research and validation are conducted.
- The work might be expanded to cover additional diseases, empowering the healthcare setting to react faster in the case of an epidemic or pandemic.

Future Scope

- In accordance with the reality that our study provides significant fresh knowledge into the COVID-19 patient survival prediction, it is essential to note that the study has a number of significant limitations.
- Concerns about the ability to be generalized of the results can be expressed in relation to the dataset that was used in the present study.
- The outcomes should be confirmed with datasets of larger size and more diversity in any future research that is performed. Additionally, the integration of additional relevant clinical variables and the investigation of new machine learning techniques may improve the prediction performance of the models.

References

[1] A. Erkoreka, “The Spanish influenza pandemic in occidental Europe (1918–1920) and victim age,” Influenza *Other Respir. Viruses*, vol. 4, no. 2, pp. 81–89, 2010, doi: 10.1111/j.1750-2659.2009.00125. x.

[2] J. Li *et al.*, “Machine learning methods for predicting human-adaptive influenza a viruses based on viral nucleotide compositions,” *Mol. Biol. Evol.*, vol. 37, no. 4, pp. 1224–1236, 2020.

[3] Y. BAI and X. TAO, “Comparison of COVID-19 and influenza characteristics,” *J. Zhejiang Univ. Sci. B*, vol. 22, no. 2, pp. 87–98, Feb. 2021, doi: 10.1631/jzus.B2000479.

[4] S. S. Aljameel, I. U. Khan, N. Aslam, M. Aljabri, and E. S. Alsulmi, “Machine Learning-Based Model to Predict the Disease Severity and Outcome in COVID-19 Patients,” *Sci. Program.*, vol. 2021, p. e5587188, Apr. 2021, doi: 10.1155/2021/5587188.

[5] Q. Liu and Y. Wu, “Supervised Learning,” Jan. 2012, doi: 10.1007/978-1-4419-1428-6_451.

Contd…

[6] N. Amruthnath and T. Gupta, "A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance," in *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*, Apr. 2018, pp. 355–361. doi: 10.1109/IEA.2018.8387124.

[7] A. Hammoudeh, "A Concise Introduction to Reinforcement Learning," Feb. 2018. doi: 10.13140/RG.2.2.31027.53285.

[8] A. Abraham Iorkaa, M. Barma, and H. Muazu, "Machine Learning Techniques, methods and Algorithms: Conceptual and Practical Insights," *Int. J. Eng. Res. Appl.*, vol. 11, pp. 55–64, Aug. 2021, doi: 10.9790/9622-1108025564.

[9] C. Hu *et al.*, "Early prediction of mortality risk among patients with severe COVID-19, using machine learning," *Int. J. Epidemiol.*, vol. 49, no. 6, pp. 1918–1929, Jan. 2021, doi: 10.1093/ije/dyaa171.

[10] X.-Y. Zhao *et al.*, "Clinical characteristics of patients with 2019 coronavirus disease in a non-Wuhan area of Hubei Province, China: a retrospective study," *BMC Infect. Dis.*, vol. 20, no. 1, p. 311, Apr. 2020, doi: 10.1186/s12879-020-05010-w.

[11] Z. Xiao, "COVID 19 Mortality Rate Prediction based on Machine Learning Methods," in *2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*, IEEE, 2021, pp. 169–177.

[12] D. Ma, S. Aaa, A. Mh, and U. P, "Automated Triage System for Intensive Care Admissions during the COVID-19 Pandemic Using Hybrid XGBoost-AHP Approach," *Sensors*, vol. 21, no. 19, Sep. 2021, doi: 10.3390/s21196379.

Prepared Manuscript

“Predicting COVID-19 Patient Survival using Machine Learning Models and Clinical Features”


Thank You!