International Journal of Computer Engineering in Research Trends
Multidisciplinary, Open Access, Peer-Reviewed and fully refereed
Research Paper Volume-10, Issue-4 ,2023 Regular Edition E-ISSN: 2349-7084

Predict Admission of Confirmed COVID-19 Cases to ICU

K. Shanmukh Akul [1], P.Y.R. Pavani [2] , A. Pradnesh [3], K. Charan Reddy [4],Jagadeesh Gopal [5]

[1,2,3,4] *Integrated M.Tech Dept. of Computer Science & Engineering, Vellore Institute of Technology University, Vellore*
[5] *Dept. of Computer Science & Engineering, Vellore Institute of Technology University, Vellore*

e-mail: katkamshanmukh.akul2019@vitstudent.ac.in, yamini.ramapavani2019@vitstudent.ac.in, pradnesh.a2019@vitstudent.ac.in, Kuppamcharan.reddy2019@vitstudent.ac.in, gjagadeesh@vit.ac.in

**Corresponding Author: katkamshanmukh.akul2019@vitstudent.ac.in*

https://doi.org/10.22362/ijcert/2023/v10/i04/v10i0409

Received: 25/03/2023, Revised: 19/04/2023, Accepted: 25/04/2023 Published:10/05/2023

Abstract: - Health systems all throughout the world have been impacted by the most recent COVID-19 pandemic. Critically sick patients have received crucial care in intensive care units (ICUs), in particular. The quick spread of the virus has increased admissions, but this has also created a number of issues for ICU wards, including a lack of ICU beds, a staff that is overworked caring for patients, and a lack of medical resources to treat everyone in hospitals. These problems may have had a direct impact on a patient's survival by lowering the quality of healthcare services offered. The project's goal is to predict the admission of covid-19 patients to ICU because we already have a shortage of beds in ICU to treat severely affected patients. This application assists hospitals in admitting critical patients only to ICU by analyzing their reports, which include various attributes collected from the patients such as temperature difference, age, blood pressure, heart rate, respiratory rate, oxygen saturation, and a few other attributes. The accurate prediction is challenging task so we use possible machine learning techniques such as logistic regression, Gaussian Naïve Bayes, SGD classifier (Stochastic gradient descent) and XGB Regressor(Extreme Gradient Boosting) and compare the performances of individual model based upon the metrics such as accuracy, precision, ROC curve, F1 score and others to implement the application with better model.

Index Terms- Gaussian Naïve Bayes, logistic regression, pandemic, SGD classifier, virus, XGB repressor

1. Introduction

When the World Health Organization declared Covid 19 a pandemic, everyone was urged to seek personal protection**.** The number of patients infected by this virus is growing by the day.Many countries have faced numerous challenges in trying to keep their health systems responsive and capable of providing essential health services. There were numerous issues raised during the treatment of critically ill patients. To avoid overburdening staff and insufficient medical resources in the event of an upcoming virus, we must improve the ICU management plan by taking into account the true patient severity. To avoid the time-consuming process and incorrect decisions made by doctors on the spot, we developed a prediction system based on training the severe covid patient data and testing the new patient data to determine the severity. The prediction system that determines whether the patient needs to be admitted to the ICU needs to be more accurate, so we investigated all available machine learning techniques that provide accurate results in less time and built a prediction system for admission of confirmed COVID-19 patients to the ICU. In this project, we aim to improve the ICU management plan by using the best machine

learning techniques to predict whether a person needs to be admitted to the ICU. This allows hospitals to treat the right person in less time. As a result, doctors in hospitals do not need to spend more time analyzing patient records and do not need to waste ICU beds.

2. Methodology

Proceed with the following steps to obtain accurate results of project:

1) Download KaggleSirioLibanesICUPrediction dataset.
2) Understand the attributes of the datasets and load.
3) Preprocess the data.
4) Train using binary classification models
    - Logistic regression
    - Gaussian Naïve Bayes
    - SGD classifier (Stochastic gradient descent)
    - XGB regressor (Extreme Gradient Boosting)
5) Predict the results using the above machine learning techniques.
6) Compare the performances of individual model and find the accurate results.

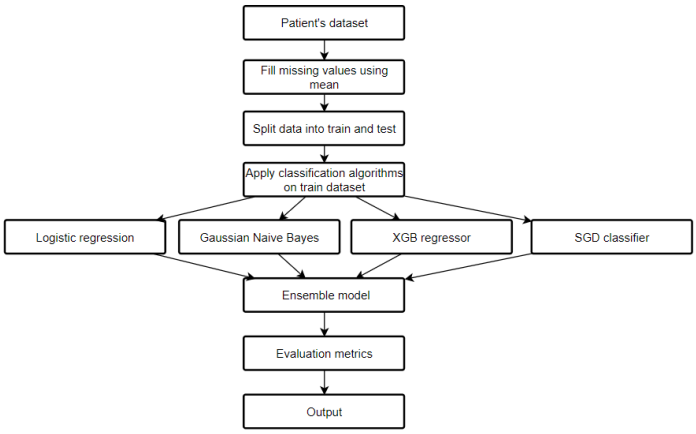

Figure 1. Flow model of the proposed work

2.1 Procedure

1. Import necessary packages and load the data set. Necessary packages or modules used:
    - Pandas: pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.
    - Numpy: Provides a huge boost in the mathematical fields with matrices and arrays
    - Matplotlib: Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.
    - Scikit learn(sklearn): Scikit-learn is a free software machine learning library for the Python programming language.
    - Stratified k fold: Stratified K-Folds cross-validator. Provides train/test indices to split data in train/test sets. This cross-validation object is a variation of KFold that returns stratified folds. The folds are made by preserving the percentage of samples for each class.

    Stratification is the process of rearranging the data as to ensure each fold is a good representative of the whole. For example in a binary classification problem where each class comprises 50% of the data, it is best to arrange the data such that in every fold, each class comprises around half the instances.

2. Understand the data set with metadata.
    Available data:
    - Patient demographic information
    - Patient previous grouped diseases
    - Blood results
    - Vital signs
    - Blood gases

    In total there are 42 features, expanded to the mean, max, min, diff and relative diff.
    The dataset used have 1925 total rows and 231 columns.

3. Clean the null values in the data using mean.
4. Split the data set into train and test data.
5. Import the model regression and train model with given data.
6. Modulate data using k-fold techniques and train the models separately (each type of regression model is trained separately two times once with straight data and again with modified data.)
7. Perform evaluation metrics for each model (accuracy, confusion matrix and roc curve).
8. Create an ensemble model using the regression types and train the ensemble model.
9. Conduct evaluation metrics for the ensemble model.

3. Implementation

Here comes the most crucial step for yourresearch publication. Ensure the drafted journal is critically reviewed by your peers or any subject matter experts. Always try to get maximum review comments even if you are well confident about your paper. We load the dataset available in Kaggle from google driveWe understand the data and the attributes.Then we count null value of each column in the dataset.

```
# Data Preparation
df['AGE_PERCENTIL'] =
df['AGE_PERCENTIL'].str.replace('Above ',
```


").str.extract(r'(.+?)th')
df['**WINDOW**'] = df['**WINDOW**'].str.replace('**ABOVE_12**', '**12-more**').str.extract(**r'(.+?)-**')

Data Preparation for check
df1['**AGE_PERCENTIL**'] = df1['**AGE_PERCENTIL**'].str.replace('**Above** ', '').str.extract(**r'(.+?)th**')
df1['**WINDOW**'] = df1['**WINDOW**'].str.replace('**ABOVE_12**', '**12-more**').str.extract(**r'(.+?)-**')

Missingness as features
df['**row_missingness**'] = df.isnull().sum(axis=1)

Missingness as features for check
df1['**row_missingness**'] = df1.isnull().sum(axis=1)
from sklearn.impute import SimpleImputer

We now clear the null values from the dataset. Since cleaning of the dataset is very important for any dataset that helps model to train in a better way.First, change the column values in string type to numerical type. Then, fill null values with mean values of the column. This can be done using imputer method in sklearn module. Then, we check the null values in the dataset. We can see that there are no null values and each column is of datatype integer in below fig.

Dividing dataset into train and testing parts. We divide data in 7:3 ratio for train and test data using train-test method in sklearn module.

target = ["**ICU**"]
un = ["**row_missingness**"]
numericals = list(set(imputed_data.columns.values) - set(target) - set(un))
numericals1 = list(set(imputed_data1.columns.values) - set(target) - set(un))

Train the binary classification models now. We are using 4 algorithms to train the data.

3.1 Logistic regression

We create model based on logistic regression and we train the model with training data. Then we evaluate the model using metrics. Accuracy of the model is reported as 83% visualizing the ROC curve of the model using matplotlib library. For visualizing ROC curve two parameters are required. Model predictions of test data part and original test values.

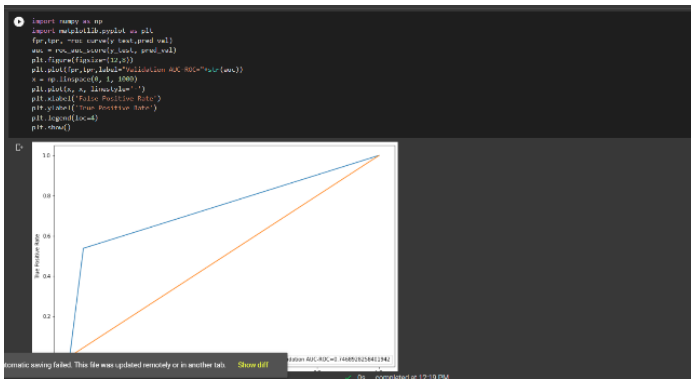

Figure 2: ROC curve

Logistic regression with stratified k fold

The accuracy of each fold is shown and the final accuracy of the model can be taken as the mean of array.
Evaluation metrics

Figure 3: Logistic regression with stratified k fold

3.2 Gaussian NB

We create a model based on Gaussian NB and we train the model with training data. We evaluate the model using metrics. Accuracy of model is reported as 77%

ROC curve:

Figure 4.Gaussian NB

Gaussian NB with stratified k-fold

Accuracy and confusion matrix as are follows:

Figure 5 : Gaussian NB with stratified k-fold

3.3 SGD classifier

We create a model based on SGD classifier and we train the model with training data. We evaluate the model using metrics.

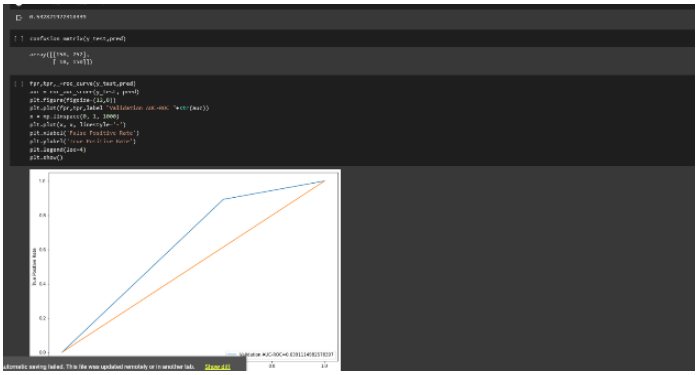

Figure 6 : SGD classifier

SGD classifier with stratified k-fold

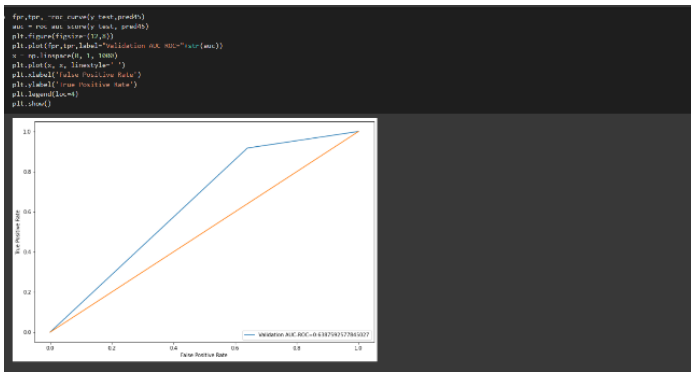

Figure 7: SGD classifier with stratified k-fold

3.4 XGB regressor

We create a model based on XGB regressor and we train the model with training data.
Accuracy of the model is reported as 89%

ROC curve:

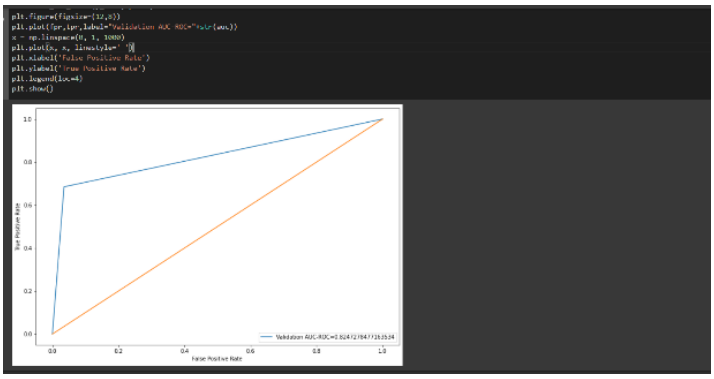

Figure 8 : ROC curve

XGB regressor with stratified k-fold

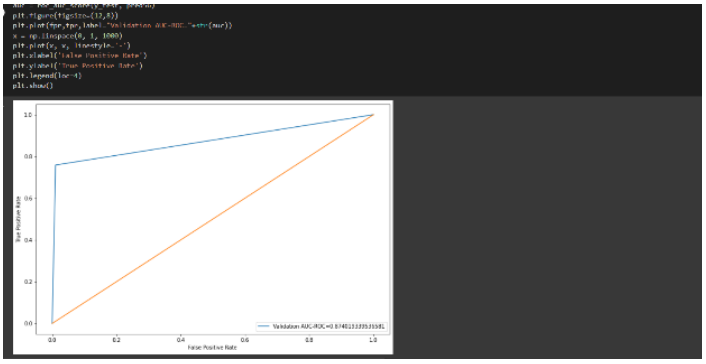

Figure9 : XGB regressor with stratified k-fold

Ensemble model

We can essemble all models into one stack model and we can train model by using the dataset. Create and run essemble model.

We can see the comparison of individual models performance and stacked model by box plot and bar plot.

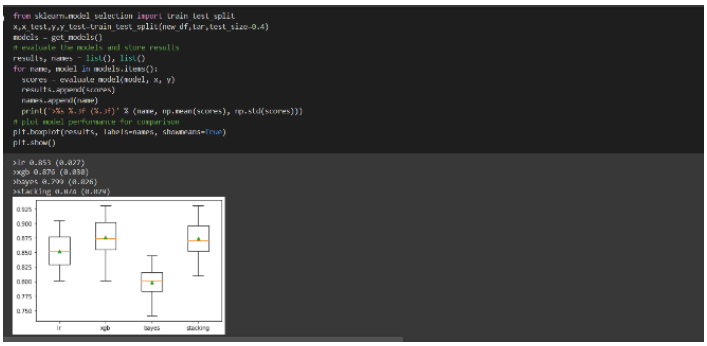

Figure 10. stack model

Figure 11: Final results

We implemented logistic regression to make interface.

We give csv file as input and predict if the person is 1(admitted to ICU) or 0(not need of admission to ICU)

4. Conclusion

In conclusion, predicting the admission of confirmed COVID-19 cases to the ICU is a critical task that can aid in the effective management of resources and ultimately improve patient outcomes. With the use of advanced machine learning algorithms and predictive models, healthcare providers can leverage data from various sources, including patient demographics, clinical characteristics, and laboratory findings, to identify patients at higher risk of ICU admission. This information can help healthcare providers prioritize resources and provide more targeted interventions to patients who are most in need. However, it is important to note that these models are not infallible, and healthcare providers should use their clinical judgment and expertise in conjunction with these predictive tools. Additionally, ongoing research and refinement of these models are necessary to ensure their accuracy and applicability across different populations and healthcare settings. Ultimately, the use of predictive models can enhance the quality of care for COVID-19 patients and help mitigate the impact of this pandemic on healthcare systems worldwide.

References

[1] M. Frid-Adar, R. Amer, O. Gozes, J. Nassar and H. Greenspan, "COVID-19 in CXR: From Detection and Severity Scoring to Patient Disease Monitoring," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 6, pp. 1892-1903, June 2021, doi: 10.1109/JBHI.2021.3069169.

[2] C. Li et al., "Classification of Severe and Critical Covid-19 Using Deep Learning and Radiomics," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 12, pp. 3585-3594, Dec. 2020, doi: 10.1109/JBHI.2020.3036722.

[3] N. Darapaneni et al., "A Machine Learning Approach to Predicting Covid-19 Cases Amongst Suspected Cases and Their Category of Admission," 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), RUPNAGAR, India, 2020, pp. 375-380, doi: 10.1109/ICIIS51140.2020.9342658.

[4] L. Famiglini, G. Bini, A. Carobene, A. Campagner and F. Cabitza, "Prediction of ICU admission for COVID-19 patients: a Machine Learning approach based on Complete Blood Count data," 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), Aveiro, Portugal, 2021, pp. 160-165, doi: 10.1109/CBMS52027.2021.00065.

[5] V. Bhadana, A. S. Jalal and P. Pathak, "A Comparative Study of Machine Learning Models for COVID-19 prediction in India," 2020 IEEE 4th Conference on Information & Communication Technology (CICT), Chennai, India, 2020, pp. 1-7, doi: 10.1109/CICT51604.2020.9312112.

[6] J. Rodríguez et al., "A Covid-19 Patient Severity Stratification using a 3D Convolutional Strategy on CT-Scans," 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 2021, pp. 1665-1668, doi: 10.1109/ISBI48211.2021.9434154.

[7] N. Darapaneni et al., "COVID 19 Severity of Pneumonia Analysis Using Chest X Rays," 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), RUPNAGAR, India, 2020, pp. 381-386, doi: 10.1109/ICIIS51140.2020.9342702.

[8] R. Y. Wang, T. Q. Guo, L. G. Li, J. Y. Jiao and L. Y. Wang, "Predictions of COVID-19 Infection Severity Based on Co-associations between the SNPs of Comorbid Diseases and COVID-19 through Machine Learning of Genetic Data," 2020 IEEE 8th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 2020, pp. 92-96, doi: 10.1109/ICCSNT50940.2020.9304990.

[9] T. Dan et al., "Machine Learning to Predict ICU Admission, ICU Mortality and Survivors' Length of Stay among COVID-19 Patients: Toward Optimal Allocation of ICU Resources," 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Korea (South), 2020, pp. 555-561, doi: 10.1109/BIBM49941.2020.9313292.

[10] A. Dhadge and G. Tilekar, "Severity Monitoring Device for COVID-19 Positive Patients," 2020 3rd International Conference on Control and Robots (ICCR), Tokyo, Japan, 2020, pp. 25-29, doi: 10.1109/ICCR51572.2020.9344386

