A Machine Learning Method For Forecasting The Duration Of Of Hospital Stay Of COVID-19 Patient

Akshat Singamsetty	Gummadi Varun Tej Gowd	J.J.Jayakanth
Department of Computational Intelligence	Department of Computational Intelligence	Assistant Professor Department of Computational Intelligencce
SRM Institute of Science and Technology,Kattankulathur, India	SRM Institute of Science and Technology,Kattankulathur, India	SRM Institute of Science and Technology,Kattankulathur, India

ABSTRACT: Longer the emergency department stays are the result of the COVID-19 pandemic's is danger to global public health and the rise in COVID-19 cases. This study's objectives are to create a trustworthy prediction model for COVID-19 patients in emergency rooms and to pinpoint clinical variables like age and comorbidities linked to a "4-hour target" for LOS. All COVID-19 patients who attended an urban, multicultural hospital in Detroit between March 2020 and December 2020 had their data gathered. The different machine learning models were created: logistic regression , gradient boosting , decision tree , and random forest, to predict the EU (LOS) Length of stay of COVID-19 patients within or beyond 4 hours. Information from 3,301 COVID-19 patients were analysed, together with information on their time spent in the emergency unit (EU) and 16 important clinical variables. The EU duration of stay was predicted using four machine learning models, including logistic regression, gradient boosting, decision trees, and random forests. The gradient boosting (GB) model performed the best, with an F1-score is 0.88 and an accuracy is 85%. In the research, important predictors of prolonged EU stays for COVID-19 patients were patient population characteristics, comorbidities, and operational EU information. This predictive model can offer patients up-to-date projections of EU length of stay and assist EUs and hospitals with resource allocation and decision-making.

Keywords – The topics being discussed are COVID-19, hospital LOS, the target, emergency departments (EDs), and machine learning methods

INTRODUCTION

Due to the Corona virus pandemic, hospitals around the world are dealing with a challenging challenge, needing more staff and resources to handle the complex care requirements and ensure patient safety. The virus has increased the number of people with confirmed or suspected infections, placing tremendous strain on emergency rooms. As a result, hospitals are struggling with a staffing and resource shortfall, necessitating the need to discover strategies to enhance patient management and care.As a result of this pandemic, health care systems in various countries have reported increased workloads and surges in patient numbers. Congestion in the EU as a result has a detrimental effect on patient outcomes and adds to the workload of medical professionals.

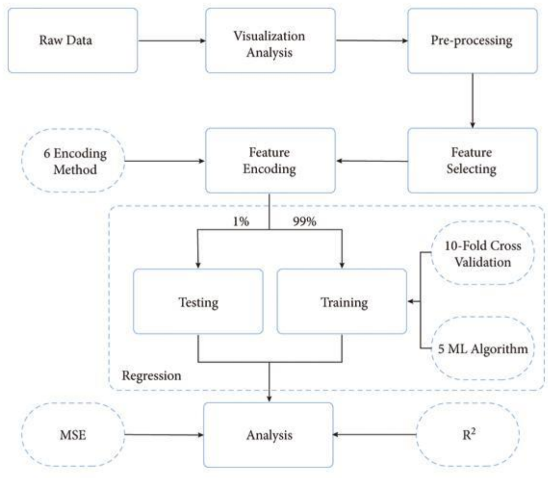

Fig.1: Example figure

The length of stay (LOS) of EU patients prior to the COVID-19 pandemic has been studied using a variety of statistical models, including decision trees, multiple linear regression, logistic regression, and increased failure time models In order to decipher complicated issues and pinpoint variables that affect COVID-19 EU patients' survival, machine learning algorithms have the capacity to take a wider variety of patient records and hospital data into account. The (LOS) Length of stay of COVID-19 EU patients has not yet been predicted by any study using these sorts of data together. In this work, a model that successfully predicted the LOS of COVID-19 EU patients through the application of four machine learning methods, including logistic regression, gradient boosting, were developed.

1. LITERATURE REVIEW

Predicting the length of hospital stay for COVID-19 patients using predictive modelling.

Due to the Corona virus pandemic's high transmission and transferable rates, ambiguous clinical patterns, absence of vaccinations and approved drug treatments, and lengthy incubation periods, healthcare organisations have encountered several difficulties [7, 8]. Given the lack of resources and the worn-out state of the healthcare workforce, hospitals have begun employing predictive algorithms to make well-informed judgements and effectively distribute resources like beds, oxygen generators, and personnel [9, 10]. This study uses data-driven machine learning methods to forecast the length of stay (LOS) of Corona virus patients in this situation. The study employed a single-center, cross-sectional design and took into account patient demographics, comorbidities, main diagnoses, and temporal factors in addition to hospital-level variables. The study included five phases, including feature selection, data description, and preprocessing.

Expecting emergency department patients to be admitted to hospitals

In the current study, the effectiveness of different Machine learning methods to estimate emergency department utilising a variety of readily available laboratory data was considered. A GaussianNB model beat previous models using the 10-fold cross-validation method in terms of the area under that is receiving the operating characteristic curve, and it showed good promise for helping in hospital admission prediction. This model is one of the scikit-learn library's most often used categorization methods.The research's shortcomings were the little amount of time that was actually monitored, the fact that all patient data came from a single study site, and the poor explainability of the recommended ML model.The main finding is that physicians may become familiar with machine learning techniques by showing them, in review the statistics of such methods and letting them consider the possibility of using the current data to compare with the admission rates over time and on the end to which current hospitals policies might be vulnerable to change more likely to have s.

Aiming for accurate forecasting of patient stay times in emergency rooms.

Data-based applications now have a wealth of options thanks to machine learning (ML) technology's quick growth, especially in the healthcare industry. Various statistical and machine learning techniques have recently been introduced to increase the precision of forecasting EU activities.The approaches described are only a few examples of these techniques. Machine learning has several uses in medicine, including enhancing ECG signal classification and arrhythmia analysis and internal medicine inpatient mortality prediction.The advancement of machine learning technology for healthcare has created a wide range of new opportunities, including better patient care and the creation of clinical decision support systems. Back-propagation neural network models are being used to forecast patient length of stay (LOS) in surgical services, taking into consideration variables like age, gender, and hospitalization data. This would assist medical teams in providing patients with individualised treatment. Similar research was conducted by Gentimis et al. (2017) on the application of neural network-based models to forecast hospital LOS utilising information on patient admission, discharge, and transfer as well as medical and laboratory data.Additionally, Pendharkar and Khurana (2014) predicted patient LOS In Pennsylvania Federal and Specialty hospitals, machine learning methods like regression trees, chi-square and automatic interaction detection

A machine learning-based algorithm has been created to forecast how long Covid-19 patients will need to spend in the critical care unit.

The availability of Intensive Care Unit (ICU) beds was one area where the Corona virus pandemic had a high impact on healthcare systems worldwide. The scientific community has thus been more interested in creating practical solutions to address the problem, such as establishing forecasting models for ICU admission rates, LOS , and discharge dates. 41 papers were analysed as part of our research to better understand this subject. To investigate the potential of machine learning (ML), we mainly concentrated on recent publications published in credible journals. Global healthcare systems have been significantly impacted by the COVID-19 epidemic, notably in terms of the availability of ICU beds. Predictive models for ICU have seen increased interest in order to address this problem. The average length of stay for Coronavirus patients in the intensive care unit, including censored patients and discharged cases, has been previously assessed using several estimate methodologies. Four predicting methods took place in this work using Machine learning classifiers such as Random Forest, gradient boost, extreme gradient booster. On a dataset, the models were trained and tested, and various tests were used to compare how well they performed. The objective was to choose the most effective classifier for creating a prediction scenario that may have an effect on public health. Previous studies have shown that the most often applied machine learning (ML) methods for predicting mortality, severity, and ICU stay duration are Random Forest and XGBoost. The findings of this study may aid healthcare professionals in identifying people who require a six-hour hospital stay. The Boruta algorithm was used to apply feature selection in order to decrease the number of features. Yet, it turned out that employing all of the features increased accuracy.

3. METHODOLOGY

Previous studies on the factors influencing EU LOS have used a variety of models, including decision trees, multiple linear regression, logistic regression, and increasedfailure time models. These models might not be able to fully reflect the complexity of the issue or pinpoint every factor influencing the Length of stay of Corona virus patients in the EU. Machine learning algorithms, on the other hand, may take into account a greater variety of characteristics and permutations, such as medical records and hospital data, to properly forecast the LOS. We are aware of no other work that has used these patient and operational data from the EU to forecast the EU LOS of COVID-19 patients.

Disadvantages:

Our study's goal was to use the different machine learning methods like logistic regression, decision trees, and random forest methods to build a prediction model for Corona virus patients' EU length of stay (LOS). To assure the model's correctness, it underwent many data processing steps of evaluation.

Advantages:

1. enhancing hospital and emergency department resource allocation, and letting patients know about better EU LOS forecasts.

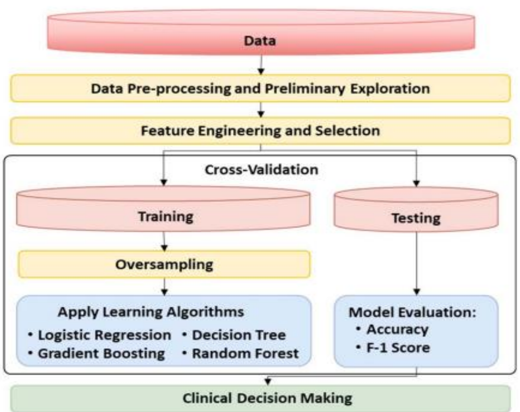

Fig.2: System architecture

MODULES:

We developed the modules indicated below to complete the aforementioned project.

- Import the packages: using this module we can import all packages
- Explore the dataset – IDS Anomaly Data:Using this method we will upload dataset
- Data processing: Using this method we will read data for processing

  - pandas dataframe

  - dropping the unwanted column
- Visualization using seaborn & matplotlib: Using this module will get graphical representation of information and data.

  · Label Encoding to convert object

  · Feature Selection

  · SMOTE Oversampling
- Splitting the data to train and test: Using this module will divide dataset into train & test for processing
- Building the model: Using this module we will build all algorithms

  - Gradient Boosting

  - Random Forest

  - Decision Tree

  - Logistic Regression

  - XGBoost

  Training the model: Using this module algorithms trained for processing & prediction.Building the model either with Voting & Random Forest Classifier since it gives better score comparing with Other Models
- Flask Framework with Sqlite for signup and signin: Using this module user will get register & login importing the packages
- User gives input as Advance Values : Uisng this method user gives input for predicting
- Trained model is used for prediction: Using this module predicted result displayed

4. IMPLEMENTATION

ALGORITHMS:

Random Forest: Popular supervised machine-learning techniques like this approach are used to address classification and regression issues. Using the mean for regression and the majority vote for classification, it builds decision trees based on a large number of samples.

Decision Tree:To decide whether a node should be divided into sub-nodes, decision trees employ a variety of strategies. The decision tree's homogeneity is enhanced when the sub-nodes appear. As a result, the node's purity with regard to the desired variable improves.

Logistic Regression: A statistical analysis technique called logistic regression can be used to forecast binary outcomes, such as "yes" or "no." In this method, the connection between a dependent variable and an independent variables is investigated. The dependent variable is predicted using the model, which is built by examining the connection between these factors.

XGBoost: Gradient-boosted decision trees (GBDT), a type of machine learning system, are used by the Extreme Gradient Boosting (XGBoost) method. This system can handle parallel tree boosting because of its capacity for scalability and dispersion. It is regarded as the best machine-learning tool for solving regression, classification, and ranking problems.

Gradient boosting:Gradient boosting methods is a machine learning methods that is used, among other things, in regression and classification applications. A prediction model is created by combining a number of less effective prediction models, typically in the form of decision trees.

5. EXPERIMENTAL RESULTS

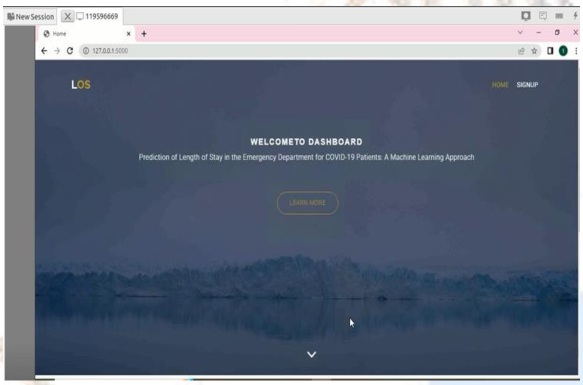

Fig.3: The Home screen

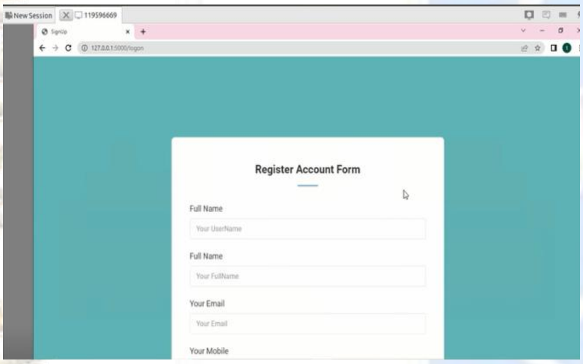

Fig.4: The User registration

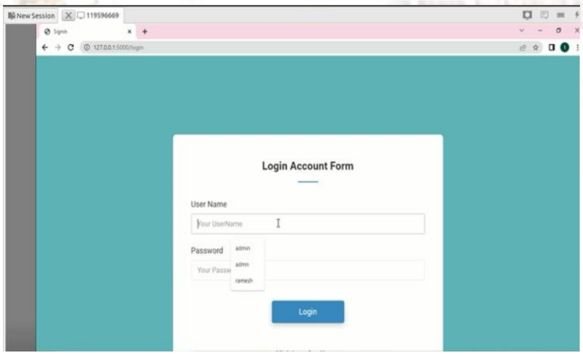

Fig.5:The user login

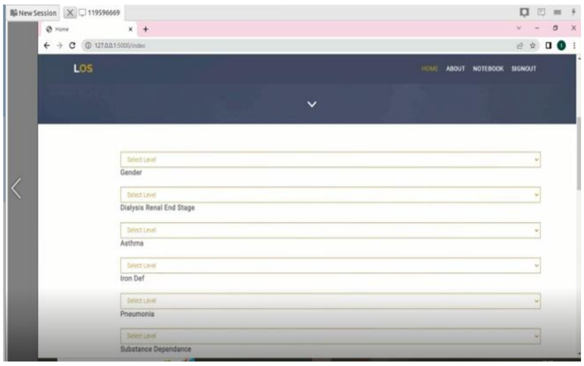

Fig.6: The Main screen

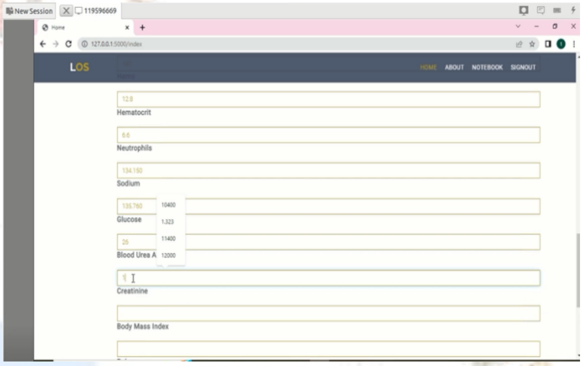

Fig.7: The User input

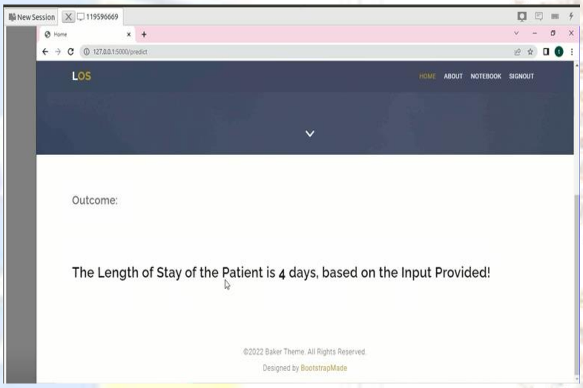

Fig.8: The Final Prediction result

6. CONCLUSION

The study's conclusions centred on certain traits of COVID-19 patients while they were being treated in hospitals, notably in the medical and emergency departments. Based on important factors connected with extended stays in Corona virus patients, the research discovered Four predicting methods were created utilising the identified characteristics to forecast the Emergency Unit Lenght of stay of COVID-19 patients. The results of this study, when combined with the framework, have the capability to be a useful decision-making tool for enhancing the provision of healthcare services, optimising resource allocation, and helping clinicians provide patients with the right treatments, particularly in lowering prolonged stays. Although these models were developed using data and medical data from Hospital, they could be modified to forecast the length of stay for Corona virus patients.

REFERENCES

[1] E. Walters, S. Najmabadi, and E. Platoff, ‘‘Texas hospitals are running out of drugs, beds, ventilators and even staff,’’ Texas Tribune, Austin, TX, USA, Tech. Rep., 2020.

[2] B. C. Sun, R. Y. Hsia, R. E. Weiss, D. Zingmond, L.-J. Liang, W. Han, H. McCreath, and S. M. Asch, ‘‘Effect of emergency department crowding on outcomes of admitted patients,’’ Ann. Emergency Med., vol. 61, no. 6, pp. 605–611, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S019606441201699X

[3] A. Guttmann, M. J. Schull, M. J. Vermeulen, and T. A. Stukel, ‘‘Association between waiting times and short term mortality and hospital admission after departure from emergency department: Population based cohort study from Ontario, Canada,’’ Brit. Med. J., vol. 342, p. d2983, Jun. 2011.

[4] U. Hwang, M. L. McCarthy, D. Aronsky, B. Asplin, P. W. Crane, C. K. Craven, S. K. Epstein, C. Fee, D. A. Handel, J. M. Pines, N. K. Rathlev, R. W. Schafermeyer, F. L. Zwemer, Jr., and S. L. Bernstein, ‘‘Measures of crowding in the emergency department: A systematic review,’’ Academic Emergency Med., vol. 18, no. 5, pp. 527–538, May 2011.

[5] N. R. Hoot and D. Aronsky, ‘‘Systematic review of emergency department crowding: Causes, effects, and solutions,’’ Ann. Emerg. Med., vol. 52, no. 2, pp. 126–136, 2008, doi: 10.1016/j.annemergmed.2008.03.014.