International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 6 Volume No.5 November – 2021 DOI: 10.46647/ijetms.2021.v05i06.002
ISSN: 2581-4621

A Novel Approach to Predict Hospital Emergency

A Sai Rekha[1], Kishore K[2]

[1]*M.Tech Student, Dept. of CSE, Dr.K.V.Subba Reddy College of Engineering for Women, Kurnool, A.P*

[2]*Assistant Professor, Dept. of CSE, Dr.K.V.Subba Reddy College of Engineering for Women, Kurnool, A.P*

[1]sairekha0503@gmail.com
[2]kishore.0331@gmail.com

***Abstract*— One of the most important tasks in the Emergency Department (ED) is to promptly identify the patients who will benefit from hospital admission. Machine Learning (ML) techniques show promise as diagnostic aids in healthcare. Material and methods: We investigated the following features seeking to investigate their performance in predicting hospital admission: serum levels of Urea, Complete Blood Count with differential, Activated Partial Thromboplastin Time, D Dimer, International Normalized Ratio, age, gender, triage disposition to ED unit and ambulance utilization. A total of 3,204 ED visits were analyzed. Results: The proposed algorithms generated models which demonstrated acceptable performance in predicting hospital admission of ED patients.**

The main advantages of this tool include easy access, availability, yes/no result, and low cost. The clinical implications of our approach might facilitate a shift from traditional clinical decision-making to a more sophisticated model. Conclusion: Developing robust prognostic models with the utilization of common biomarkers is a project that might shape the future of emergency medicine. Our findings warrant confirmation with implementation in pragmatic ED trials.

Patients boarding in the Emergency Department can contribute to overcrowding, leading to longer waiting times and patients leaving without being seen or completing their treatment. The early identification of potential admissions could act as an additional decision support tool to alert clinicians that a patient needs to be reviewed for admission and would also be of benefit to bed managers in advance bed planning for the patient. We aim to create a low-dimensional model predicting admissions early from the pediatrics Emergency Department.

***Keywords*— Data mining, emergency department, hospitals, machine learning, predictive models.**

I. INTRODUCTION

Predicting admissions early in the patient's journey through the pediatrics Emergency Department (ED) has potential to improve the patient flow system through both the ED and hospital. One of the influential factors contributing to overcrowding in the pediatrics ED is the presence of patients boarding in the treatment area that require admission but cannot leave the ED due to lack of bed capacity in the hospital. As the volume of patients arriving increases, space, resources, and clinical needs may become an issue as a result of patients boarding in the treatment area, increasing the waiting time for other patients in the waiting room and can cause less acute patients to leave without being seen or before the completion of their treatment. Early admission prediction would provide advance notice to both ED clinicians and bed managers facilitating decision support and bed planning.

The benefit of using machine learning algorithms to predict admissions was realized in some of the first studies that compared clinical judgment to that of machine learning algorithms with many researchers acknowledging that clinical judgment alone, at an early stage, is not enough to accurately predict an outcome of admission. A review of the literature has revealed many diverse studies proposing a solution to the question of whether admissions can be predicted from the ED using machine learning algorithms. Some that focus on admission prediction for specific cohorts of patients such as acute bronchiolitis and asthma and others investigating the use of natural language processing to extract valuable information from unstructured text. A few researchers have concentrated on early prediction or progressive time approaches, adding extra information to the model as the patient moves through the ED. There have also been comparisons made between the different machine learning algorithms, with many outperforming the traditional logistic regression classifier. The development of tools using minimal predictors to calculate risk of admission scores in some studies has underlined the importance of identifying strong predictors for model development.

A review of 26 studies that looked at predicting admissions from the ED provides valuable insight into the types and significance of predictors used. The most frequently used predictors were age, sex, triage category, presenting complaint/symptoms, and arrival mode. Apart from sex these were also reported as some of the most influential for predicting admission, particularly at an early stage. To further increase model performance numerous researchers included significant predictors such as vitals, pain scores, anthropometrics, medication, radiology, and laboratory tests ordered. For one pediatric study that created models after 0, 10, 30 and 60 min, the inclusion of these types of predictors resulted in an Area Under the Curve (AUC) of 0.789 for 0 min up to an outstanding discrimination value of 0.913 at 60 min upon evaluation.

II.SYSTEM ANALYSIS

This study will follow the data mining methodology, Cross Industry Standard Process for Data Mining (CRISP-DM) consisting of 6 key business understanding, data understanding, data preparation, modelling, evaluation, and deployment. Data extraction and transformation will be performed using Microsoft SQL Server Management Studio, with subsequent data

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 6 Volume No.5 November – 2021 DOI: 10.46647/ijetms.2021.v05i06.002
ISSN: 2581-4621

preparation, modelling, and evaluation to be carried out using R Studio Version 1.1.456. From 3 different machine learning algorithms and 5 sampling techniques, 15 models will be developed. The best performing model will be selected based on the highest AUC, from which the variables of importance will also be derived and used to create a further low-dimensional model.

Data Sources and Sample Size

Data will be extracted from 3 separate information systems and will use the patient's healthcare record number as the common link. Most of the data will be retrieved from the ED information system, with the patient administration system and inpatient enquiry system providing hospital admission usage and medical history data. The study sample will consist of 2 years of data from 2017 to 2018, providing a good of representation of seasonal changes and the unique values within each variable. Based on the average attendance per year, the sample size will be ~76,000.

Study Participants and Exclusion Criteria

All attendances to one acute pediatric ED in the Republic of Ireland will be included. Visits will be excluded for the following:

1. Patients over 18 years of age.

2. Visits where the patient left without being seen or left before completion of treatment.

3. Patients returning for direct day case surgical management.

Missing data will be analysed, listwise deletion will be performed depending on the percentage of missing values and whether those values are missing at random. Otherwise the most appropriate principled method to handle missing data will be applied. These methods may include multiple imputation, expectation-maximum algorithm or full information maximum likelihood.

Outcome and Predictors

The outcome to be predicted is “admission” or “discharge.” Patient visits with a discharge outcome of admission, transferred to another hospital for admission and died in department will be grouped into the category of “admission,” all other visit discharge outcomes will be defined as “discharge.”

Based on a review of the literature the following predictors, comprised of both numerical and categorical data types will be included in the study.

Demographics

Age, sex, and distance travelled. Distance travelled will be measured in kilometres and will be calculated from the patient's home address to the hospital site.

Registration Details

Arrival mode, referral source, registration date and time (split into weekday, month, and time), re-attendance within 7 days, presenting complaint and infection control alert.

III.SYSTEM CONSTRUCTION

Three machine learning algorithms will be used to compare performance across the 5 different training sets, resulting in the development of 15 models (Figure 1). Logistic regression which is the traditional choice of classifier for this field of study will be compared with naïve Bayes and the ensemble method, gradient boosting machine. These machine learning algorithms were selected as they can be used directly with categorical data that has not been encoded. Both logistic regression and naïve Bayes were used extensively in previous studies, with the gradient boosting machine algorithm achieving a higher AUC than other classifiers , therefore providing a good basis for comparison. The optimal tuning parameters for both the naïve bayes and the gradient boosting machine algorithms will be selected by creating a custom tuning grid and using 10-fold cross validation.

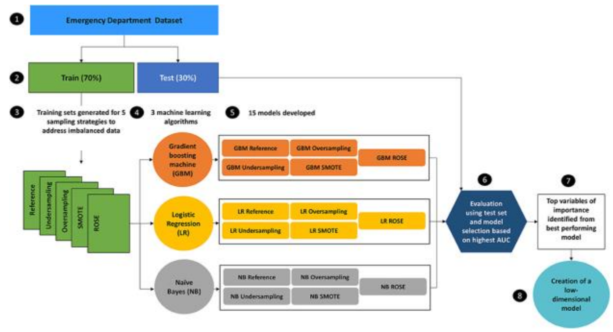

Figure 1. *Design of experiment to identify the model with the highest Area Under the Curve (AUC)*

The models will be validated and evaluated by applying the test set. Performance will be measured primarily using AUC, with specificity, sensitivity, accuracy, positive prediction value and negative prediction being produced as the secondary measurements. Confidence Intervals at 95% will be generated for each measure. When reporting these measures and to assist comparison, the specificity will be fixed at 90% to evaluate the true impact of applying the different sampling methods for imbalance at a common fixed point.

The variables of importance will be obtained from the model with the highest AUC. The calculation of relative importance of each predictor will differ depending on the machine learning algorithm and will be calculated for the optimal model only. For logistic regression, the odds ratios and regression coefficients will be produced. The a priori and conditional probabilities will be examined for naïve Bayes and the average decrease in mean squared error for the gradient boosting machine will be produced. A low-dimensional model will then be created based on the top variables of importance. The number of dimensions to be included will be determined by assessing the AUC, beginning with the top 10 variables, and reducing the number of variables according to relative importance.

IV.CONCLUSION

We propose creating a low-dimensional machine learning prediction model based on routinely collected data up to the post-triage process. From the literature review, the most common and successful predictors were obtained and used to assess which data could be included in the formation of our dataset. Not all hospital environments are at the same level of information technology maturity and therefore may also have limited data to form these datasets, with many predictors heralded as being significant in previous studies, not available to them. The approach we have taken focuses more on generalisability, by identifying significant predictors to use in a low-dimensional model. A model that will use 10 or less variables based on commonly collected data to make a prediction. In a study generalizing a model was explored, evident from this study was the low number of predictors included (6 in total), although AUC results were lower than more recent studies that included more variables, the study successfully demonstrated how a low-dimensional model could be used across different hospitals..

The three models presented in this study yield comparable, and in some cases improved performance compared to models presented in other studies. Implementation of the models as a decision support tool could help hospital decision makers to more effectively plan and manage resources based on the expected patient inflow from the ED. This could help to improve patient flow and reduce ED crowding, therefore reducing the adverse effects of ED crowding and improving patient satisfaction.

V.REFERENCES

[1] A. Cameron, K. Rodgers, A. Ireland, R. Jamdar, and G. A. McKay, ``A simple tool to predict admission at the time of triage," *Emerg. Med. J.*, vol. 32, no. 3, pp. 174_179, 2013, doi: 10.1136/emermed-2013-203200.

[2] N. Esfandiari, M. R. Babavalian, A. M. E. Moghadam, and V. K. Tabar, ``Knowledge discovery in medicine: Current issue and future trend," *Expert Syst. Appl.*, vol. 41, no. 9, pp. 4434_4463, 2014, doi: 10.1016/j.eswa.2014.01.011.

[3] H. C. Koh and G. Tan, ``Data mining applications in healthcare," *J. Healthcare Inf. Manag.*, vol. 19, no. 2, pp. 64_72, 2005.

[4] C. Baker. (2015). *Accident and Emergency Statistics, England*. [Online]

[5] DHSSPS. (2015). *Northern Ireland Hospital Statistics: Emergency Care (2014/15), Northern Ireland*. [Online]. Available:https://www.dhsspsni.gov.uk/sites/default/_les/publications/dhssps/hsemergency-

care-2014-15.pdf

[6] M.-L. Connolly. (2015). *NI Emergency Healthcare Enquiry Finds `Degrading' Cases*. Accessed: Oct. 14, 2015. [Online]. Available: http://www.bbc.co.uk/news/uk-northern-ireland-32888240

[7] (2014). *Royal Victoria Hospital: Delays `Contributed to Five*

Deaths. Accessed: Oct. 14, 2015. [Online]. Available: http:// www.bbc.co.uk/news/uk-northern-ireland-26135756

[8] J. P. Ruger, L. M. Lewis, and C. J. Richter, ``Identifying high-risk patients for triage and resource allocation in the ED," *Amer. J. Emerg. Med.*, vol. 25, pp. 794_798, Sep. 2007, doi: 10.1016/j.ajem.2007.01.014.

[9] W.-T. Lin, S.-T. Wang, T.-C. Chiang, Y.-X. Shi, W.-Y. Chen, and H.-M. Chen, ``Abnormal diagnosis of emergency department triage explored with data mining technology: An emergency department at a medical center in Taiwan taken as an example," *Expert Syst. Appl.*, vol. 37, no. 4, pp. 2733_2741, 2010, doi: 10.1016/j.eswa.2009.08.006.

[10] M. R. Baumann and T. D. Strout, ``Triage of geriatric patients in the emergency department: Validity and survival with the emergency severity index," *Ann. Emerg. Med.*, vol. 49, no. 2, pp. 234_240, 2007, doi: 10.1016/j.annemergmed.2006.04.011.

[11] D. M. Fatovich, Y. Nagree, and P. Sprivulis, ``Access block causes emergency department overcrowding and ambulance diversion in Perth, Western Australia," *Emerg. Med. J.*, vol. 22, no. 5, pp. 351_354, 2005, doi: 10.1136/emj.2004.018002.

[12] M. L. McCarthy *et al.*, ``Crowding delays treatment and lengthens emergency department length of stay, even among high-acuity patients," *Ann. Emerg. Med.*, vol. 54, no. 4, pp. 492_503, 2009, doi: 10.1016/j.annemergmed.2009.03.006.

[13] D. B. Richardson, ``Increase in patient mortality at 10 days associated with emergency department overcrowding," *Med. J. Aust.*, vol. 184, no. 5, pp. 213_216, 2006.

[14] 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.