EFFECTIVE PREDICTION OF PATIENT ADMISSION IN HOSPITAL USING DATA MINING CLASSIFICATION ANALYTICAL TOOL

Ms. N. Zahira Jahan M.C.A., M.Phil.,*,
Ms. K. Priyanka**
* Associate Professor, Department of MCA, Nandha Engineering College, (Autonomous), Erode, Tamil Nadu, India.
Email: zahirajahan1977@gmail.com
** Final MCA, Department of MCA, Nandha Engineering College, (Autonomous), Erode, Tamil Nadu, India.
Email: priyanka13022019@gmail.com

****

Abstract:

In hospitals, managing crowd is a big issue. Particularly, in emergency departments may create major negative consequences for patients. Emergency departments need to explore the use of innovative methods to improve the patient flow and prevent overcrowding. One potential method is the use of data mining using machine learning techniques to predict emergency department admissions. This project uses routinely collected administrative data and to compare contrasting machine learning algorithms in predicting the risk of admission from the emergency department. The existing system draws on this data to achieve two objectives. First one is to make a model that accurately predicts admission to hospital from the emergency department. Second is to measure the performance of common machine learning algorithms in predicting hospital admissions. This project suggests using the cases for the implementation of the model as a decision support and performance management tool. The logistic regression and decision tree models presented in this project 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 emergency department. This could help to improve patient flow and reduce emergency department crowding, therefore reducing the adverse effects of emergency department crowding and improving patient satisfaction.The models even have potential application in performance monitoring and audit by comparing predicted admissions against actual admissions. However, at the same time as the model could be used to support planning and decision making, individual level admission decisions still require clinical judgments.

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

****

I. INTRODUCTION

Medical Applications is an interdisciplinary field that develops methods and software tools for understanding biological data. As an interdisciplinary field of science, Medical Applications combines computer science, statistics, mathematics, and engineering to analyze and interpret biological data. With emerging new concepts, theories and techniques in biological analysis, huge amount of data is being collected by scientists after conducting various experiments. Though the quantity of knowledge grows exponentially, it's becomes impractical to research them manually. This is where computer science techniques intervene together with statistics, mathematics and engineering. Computational techniques are used to analyze these large amounts of data more accurately and efficiently. Hence, Medical Applications can be considered as a field of data science for solving problems in biology. Medical Applications deals with biology and biological data. Medical Applications, or computational biology, is the interdisciplinary science of interpreting biological data using information technology and computer science. The importance of this new field of inquiry will grow as we still generate and integrate large quantities of genomic, proteomic, and other data. A particular active area of research in Medical Applications is the application and development of data mining techniques topological problems. Analyzing large biological data sets requires making sense of the info by inferring structure or generalizations from the info. Examples of this sort of study include protein structure prediction, gene classification, cancer classification supported microarray data, clustering of organic phenomenon data, statistical modeling of protein-protein interaction, etc. Therefore, we see a great potential to increase the interaction between data mining and Medical Applications. Medical Applications involves the manipulation, searching and data mining of DNA sequence data. The development of techniques to store and search DNA sequences have led to widely applied advances in computer science, especially string searching algorithms, machine learning and database theory.

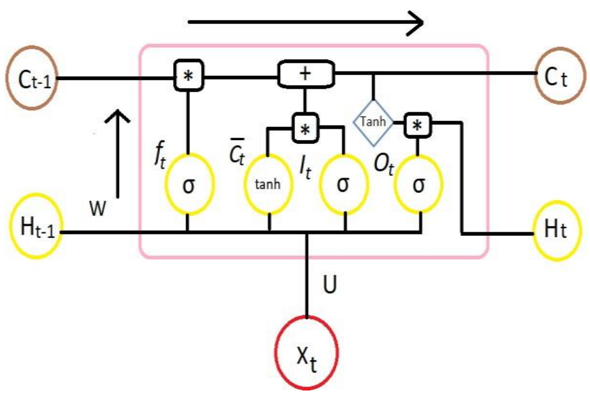
NEED FOR DATA MINING IN MEDICAL APPLICATIONS

The entire human genome, the entire set of genetic information within each human cell has now been determined. Understanding these genetic instructions promises to permit scientists to raise understand the character of diseases and their cures, to spot the mechanisms underlying biological processes like growth and ageing and to obviously track our evolution and its relationship with other species. The key obstacle lying between investigators and therefore the knowledge they seek is that the sheer volume of knowledge available. Biologists, like most natural scientists, are trained primarily to collect new information. Until recently, biology lacked the tools to research massive repositories of data like the human genome database. Luckily, the discipline of computer science has been developing methods and approaches well suited to help biologists manage and analyze the incredible amounts of data that promise to profoundly improve the human condition. Data mining is one such technology.

LONG SHORT TERM MEMORY (LSTM)

LSTM may be a quite recurrent neural network. In RNN output from the last step is fed as input within the current step. LSTM was designed by Hochreiter & Schmidhuber. It tackled the matter of long-term dependencies of RNN during which the RNN cannot predict the word stored within the LTM but can give more accurate predictions from the recent information. As the gap length increases RNN does not give efficient performance. LSTM can by default retain the knowledge for long period of your time. It is used for processing, predicting and classifying on the idea of your time series data.

Structure of LSTM



II.RELATED WORK

LaMantia et al. [1] used logistic regression to predict admissions to hospital, and ED re-attendance. They predicted admissions with moderate accuracy, but were unable to predict ED re-attendance accurately. The most important factors predicting admission were age, Emergency Severity Index (ESI) triage score, ED rate, diastolic blood pressure, and chief complaint (pg. 255). Baumann and Strout also find an association between the ESI and admission of patients aged over 65.

Boyle et al. [2] used historical data to develop forecast models of ED presentations and admissions. Model performance was evaluated using the mean absolute percentage error (MAPE), with the best attendance model achieving a MAPE of around 7%, and the best admission model achieving a MAPE of around 2% for monthly admissions. The use of historical data by itself to predict future events has the advantage of allowing forecasts further into the future, but has the disadvantage of not incorporating data captured at arrival and through triage, which may improve the accuracy of short term forecasting of admissions. Sun et al. [3] developed a logistic regression model using two years of routinely collected administrative data to predict the probability of admission at the point of triage. Risk of admission was related to age, ethnicity, arrival mode, patient acuity score, existing chronic conditions, and prior EDattendances or admission in the past three months. Although their data showed the admission of more females than males, sex was not significant in the final model. Similarly, Cameron et al. [4] developed a logistic regression model to predict the probability of admissions at triage, using two years of routine administration data collected from hospitals in Glasgow. The most important predictors in their model included 'triage category, age, National Early Warning Score, arrival by ambulance, referral source, and admission within the last year' , with an area under the curve of the receiver operating characteristic (AUC-ROC) of 0.877. Other variables including weekday, out of hours attendances, and female gender, were significant but did not have high enough odds ratios to be included in the final models.

Kim et al. [5] used routine administrative data to predict emergency admissions, also using a logistic regression model. However, their model was less accurate with an accuracy of 76% for their best model. Although these models highlight the usefulness of logistic regression in predicting ED admissions, achieved better performance using a Coxian Phase model over logistic regression model, with the former AUC-ROC of 0.89, and the latter 0.83.

B. Graham et al. [6] using Data Mining to Predict Hospital Admissions From ED models were the most accurate, both predicting just over 80% of cases correctly, with FMM (with a genetic algorithm) predicting 77.97% of cases correctly. Similarly, Peck et al. [7] developed three models to predict ED admissions using logistic regression models, naïve Bayes, and expert opinion. All three techniques were useful in predicting ED admissions. Variables in the model included age, arrival mode, emergency severity index, designation, primary complaint, and ED provider. Their logistic regression model was the most accurate in predicting ED admissions, with an AUC-ROC of 0.887. Perhaps surprisingly, this model performed better than triage nurse's opinion regarding likely admission. The use of logistic regression to predict admission was subsequently found to be generalizable to other hospitals.

Peck et al. [8] have shown that the use of the predictive models to prioritise discharge or treatment of patients can reduce the amount of time the patient spends in the ED department.

Qui et al. [9] used a relative vector machine to predict whether an ED attender would be discharged or admitted to one of three hospital words. Their model had an overall accuracy of 91.9% with an AUC of 0.825. However, the accuracy of predicting the target ward varied by ward and by the probability threshold used.

Lucini et al. [10] used eight common machine learning algorithms to predict admissions from the ED department supported features derived from text recorded on the patient's record. Six out of the eight algorithms had similar levels of performance including nu support vector machines, support vector classification, extra trees, logistic regress, random forests, and multinomial naïve bayas, with Gadabouts and a choice tree performing worst. Taking a special approach, prediction for HD dataset model.

III. METHODS

MODULE DESCRIPTION

Normalization
Feature Selection
Classification

A. NORMALIZATION PROCESS

Normalization is that the process of classify data into an associated table it also eliminates redundancy and increases the reliability which improves output of the query. To normalize a database, we divide the ED dataset into tables and establish relationships between the tables. Dataset normalization can essentially be defined because the practice of optimizing table structures. Optimization is accomplished as a result of a radical investigation of the varied pieces of knowledge which will be stored within the database, especially concentrating upon how this data is interrelated

B. FEATURE EXTRACTION

PSO feature extraction model for emergency departments dataset and applied an improve probability in many medical application such as training artificial neural networks, linear constrained function optimization, wireless network optimization, data classification, and lots of other areas where GA are often applied. Computation in NN is based on a swarm of processing elements called number of network in which each node represent a candidate solution.

The system is initialized with ED dataset of random solutions and searches for optima by updating emergency department dataset generations. The search process utilizes a mixture of deterministic and probabilistic rules that depend upon information sharing among their population members to reinforce their search processes. Emergency department's prediction system sharing mechanism in NN is considerably different. In GAs, chromosomes share information with each other, so the whole emergency department's dataset moves like one group towards a selected area. In NN, the global best routing found among the hospital is the only emergency departments dataset shared among different dataset. It is a one - way emergency department's prediction sharing mechanism. The emergency departments prediction computation time in NN is much less than in GAs because all swam particles in NN end to meet to the best solution fast.

C. CLASSIFICATION

The basic classification is based on supervised algorithms. Algorithms are applicable for the input data. Classification is completed to understand the exactly how data is being classified. The Classify Tab is additionally supported which shows the list of machine learning algorithms. These algorithms generally operate a classification algorithm and run it multiple times manipulating algorithm parameters or input file weight to extend the accuracy of the classifier.

- Random Forest
- SVM Classification

Support Vector Machine (SVM)

A sequential minimal optimization (SMO) is a learning system that uses a hypothesis space of linear functions in a high dimensional space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory. SVM uses a linear model to implement non-linear class boundaries by mapping input vectors non-linearly into a high dimensional feature space using kernels. The training emergency departments' dataset examples that are closest to the maximum margin hyper plane are called support vectors. All classification models other emergency department's dataset training examples are irrelevant for defining emergency department's prediction point the binary class boundaries. The support vectors are then used to construct emergency department's dataset model and an optimal is a linear regression emergency department's prediction function (in case of regression) in this feature space. Support vector machines are supervised emergency departments prediction learning models with associated learning algorithms that emergency departments dataset analyse data and recognize emergency departments prediction state, used for classification and regression accuracy analysis.

Neural Network Classification Model

A Neural Network Classification model (NN) may be a feed forward artificial neural network model that maps ED datasets of input file onto a group of appropriate outputs. An NN classification may be a multiple layers of nodes during a directed graph, with each layer fully connected to subsequent one. Apart from the input nodes, each node may be a neuron (or processing element) with a nonlinear activation function. NN classification ED dataset utilizes a supervised learning technique called back propagation for training ED the network. MLP may be a change of the quality linear model and may distinguish data that aren't linearly separable ED dataset process. These networks are directed acyclic graphs that allow efficient demonstration of the joint ED attribute probability distribution over a group of random attribute variables. Each vertex within the graph represents a random attribute variable, and edges represent direct correlations between the attribute variables. More precisely, the network encodes the subsequent conditional independence statements: each attribute variable is independent of its non-descendants within the graph given the state of its parents. These independencies are then exploited to scale back the amount of parameters needed to characterize a probability distribution, and to efficiently compute posterior probabilities given evidence.

Attribute based Probabilistic value for ED prediction parameters model is encoded during a set of tables, one for every variable, within the sort of local conditional distributions of a variable given its parents ED disease. Using the independence statements encoded within the network, the joint distribution is uniquely determined by these local conditional distributions. Bayesian networks ED dataset classification are factored representations of probability ED dataset distributions that generalize the naive Bayesian classifier and explicitly represent statements about independence ED prediction state.

IV.RESULTS AND DISCUSSION

As shown in table, the Random forest performs best across all performance measures. A small difference is observed the remaining two methods decision tree and gradient boosted machine. Here it's seen that call tree is performing better than the gradient boosted machine. This study gives a broad spectrum of various methods of machine learning utilized in the sector of healthcare. The prediction of the hospital admission from emergency department helps the hospital management for resource planning. This also reduces the waiting time of the patient which is carried while the triage process. For admission of a patient by the emergency département the triage process plays an important role.

V.CONCLUSION

This study involved the event and comparison of three machine learning models aimed toward predicting hospital admissions from the ED. Each model was trained using routinely collected ED data using three different data processing algorithms, namely logistic regression, decision trees and gradient boosted machines. Overall, the GBM performed the simplest in comparison to logistic regression and decision trees, but the choice tree and logistic regression also performed well. The three models presented during this study yield comparable, and in some cases improved performance compared to models presented in other studies. Implementation of the models as a choice support tool could help hospital decision makers to more effectively plan and manage resources supported the expected patient inflow from the ED. this might help to enhance patient flow and reduce ED crowding, therefore reducing the adverse effects of ED crowding and improving patient satisfaction. The models even have potential application in performance monitoring and audit by comparing predicted admissions against actual admissions. However, whilst the model might be wont to support planning and deciding, individual level admission decisions still require clinical judgement.

REFERENCES

[1] M. A. LaMantia et al., ‘‘Predicting hospital admission and returns to the emergency department for elderly patients,’’ Acad. Emerg. Med., vol. 17, no. 3, pp. 252–259, 2010, doi: 10.1111/j.1553-2712.2009.00675.x.

[2] J. Boyle et al., ‘‘Predicting emergency department admissions,’’ Emerg.Med. J., vol. 29, pp. 358–365, May 2012, doi: 10.1136/emj.2010.103531.

[3] Y. Sun, B. H. Heng, S. Y. Tay, and E. Seow, ‘‘Predicting hospital admissions at emergency department triage using routine administrative data,’’ Acad.Emerg. Med., vol. 18, no. 8, pp. 844–850, 2011, doi: 10.1111/j.1553-2712.2011.01125.x.

[4] 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.

[5] S. W. Kim, J. Y. Li, P. Hakendorf, D. J. O. Teubner, D. I. Ben-Tovim, and C. H. Thompson, ``Predicting admission of patients by their presentation to the emergency department,'' *Emerg. Med. Austral.*, vol. 26, no. 4, pp. 361_367, 2014, doi: 10.1111/1742-6723.12252.

[6] Graham et al.,”Using Data Mining to Predict Hospital Admissions From the Emergency Department”, IEE Access vol.6 , pp. 10458-10469, 2018, doi: 10.1109/ACCESS.2018.2808843.

[7] J. S. Peck et al., ‘‘Generalizability of a simple approach for predicting hospital admission from an emergency department,’’Acad. Emerg. Med., vol. 20, pp. 1156–1163, Nov. 2013, doi: 10.1111/acem.12244.

[8] J. S. Peck, J. C. Benneyan, D. J. Nightingale, and S. A. Gaehde, ``Predicting emergency department inpatient admissions to improve same-day patient _ow,'' *Acad. Emerg. Med.*, vol. 19, no. 9, pp. 1045_1054, 2012, doi: 10.1111/j.1553-2712.2012.01435.x.

[9] S. Qiu, R. B. Chinnam, A. Murat, B. Batarse, H. Neemuchwala, and W. Jordan, ``A cost sensitive inpatient bed reservation approach to reduce emergency department boarding times,'' *Health Care Manag. Sci.*, vol. 18, no. 1, pp. 67_85, 2015, doi: 10.1007/s10729-014-9283-1.

[10] F. R. Lucini *et al.*, ``Text mining approach to predict hospital admissions using early medical records from the emergency department,'' *Int. J. Med. Inf.*, vol. 100, pp. 1_8, Apr. 2017, doi: 10.1016/j.ijmedinf.2017.01.001.