ISSN: 2321-2152

# IJMECE

*International Journal of modern electronics and communication engineering*

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152 www.ijmece .com

Vol 8, Issuse.2 June 2020

# Using Data Mining to Predict Hospital Admissions From the Emergency Department

*S.Raju[1],E.Muralidhar Reddy[2],G.Shankar Rao[3],G.Harika[4],*

**ABSTRACT**

***There is a risk that people may suffer serious harm as a result of overcrowding in emergency departments (EDs). As a result, emergency clinics must explore employing innovative techniques to boost patient flow while simultaneously minimising congestion in the waiting area. One option for projecting emergency department admissions is to use data mining and machine learning technologies to anticipate ED admissions. This study, which takes use of routinely obtained administrative data (120 600 records) from two major acute hospitals in Northern Ireland, presents a comparison of two rival machine learning algorithms for predicting the risk of admission from the emergency department at the hospital. Three algorithms are used in the process of developing the prediction models: A decision tree may be divided into three types, which are as follows: Decision trees include: 1) decision trees, 2) gradient boosted machines, and 3) logistic regression, which are all types of decision trees (GBM). The GBM has an AUC-ROC D of 0:824 which was better than both the decision tree and the logistic regression model (accuracy D 80:06 percent, AUC-ROC D 0:824). In this case, the accuracy is 80:06 percent and the AUC-ROC is 0:824. In this situation, the accuracy is 80:31 percent, and the AUC-ROC is 0:859, which indicates a good fit. (0:849) (0:849) (AUC-ROC D 0:849) (accuracy D 79:94 percent). We discovered a number of factors that were connected with hospital admissions via the application of logistic regression. These considerations included hospital location, age, arrival mode, triage category, care group, and previous hospitalisation during the previous month or year, among other things. This study highlights the potential value of machine learning systems by using three fundamental machine learning algorithms to predict patient admissions. Decision support systems may be able to offer a picture of expected ED admissions at any given moment as a consequence of this study, allowing for resource planning ahead of time and avoiding patient flow bottlenecks. This research also suggests that the models described in this study may be utilised to perform comparisons between projected and actual admission rates. Generalised bivariate models (GBMs) are sufficient when interpretability is a concern; however, if accuracy is crucial, logistic regression models should be considered.***

***INDEX TERMS** Data mining, emergency department, hospitals, machine learning, predictive models.*

## I. INTRODUCTION

Overcrowding in emergency departments (EDs) may result in longer wait times, ambulance diverts, decreased staff morale, worse patient outcomes, such as greater mortality, and the cancellation of elective treatments, to name a few consequences. Surplus capacity in emergency rooms is a serious global concern [7], encouraging the development of innovative ways to alleviate the situation [4]. According to a previous research, overcrowding in emergency departments is a major worldwide problem [7]. The presence of increasing ED visits, inaccurate ED visits, the absence of alternative treatment options, a scarcity of inpatient beds, emergency department staffing shortages, and the closure of other adjacent ED departments are some of the factors that contribute to emergency department congestion, depending on the situation [1, 8]. In particular, the difficulty in transferring patients to an inpatient bed [1, 2] is one of the most significant of these issues, making it imperative for hospitals to manage patient flow and analyse inpatient bed availability and demand [4, 5]. One method that may help to reduce ED congestion and improve patient flow is the use of data mining to identify patients who are at high risk of inpatient admission and making attempts to avoid system bottlenecks [9, 10]. Inpatient bed management, staff planning, and supporting specialised work streams within the emergency department are just a few of the applications that might benefit from a model that can consistently estimate hospital admissions. Cameron and his friends have been taken into custody.

*Professor[1,2,3,4], Assistant Professor[1,2,3,4], ,Associate Professor[1,2,3,4],*
*Department of CSE Engineering,*
*Pallavi Engineering College,*
*Mail ID:srajunayak@gmail.com, Mail ID:krishna81.reddy@gmail.com,*
*Kuntloor(V),Hayathnagar(M),Hyderabad,R.R.Dist.-501505*

Patients may be more satisfied if they are informed ahead of time that they will be admitted to the hospital, according to the literature [11], which supports the use of the method. The use of data mining procedures, which entail researching and analysing data in order to extract crucial information and knowledge from which judgements may be made, may be required in order to develop such a model. [13] In order to do so, it is often essential to recognise and explain data patterns, as well as to create forecasts based on previous trends. It is the goal of this study to evaluate the usefulness of various model-building approaches, including the use of machine-learning algorithms, in order to produce models that forecast hospital admissions from community-based emergency departments in the future. In order to train and test the models, it was required to gather data from two acute hospitals in Northern Ireland, and we obtained this information via their administrative systems.

Emergency department performance in Northern Ireland has been a major source of concern for a number of years, particularly in recent years. Demand for emergency services in Northern Ireland has increased, putting a burden on the province's emergency departments, which has resulted in poor performance across the province when compared to other areas of the United Kingdom [14, 15]. Several studies, for example, have shown that no emergency department in Northern Ireland met the 4 hour wait time objective in June 2015 [15], with over 200 patients waiting more than 12 hours throughout the region to be sent to a hospital or sent home [16]. The media [16, 17] has reported that this may have a negative impact on patients at various stages of their treatment journey. [16, 17] In the emergency department, depending on their decisions made in earlier stages, patients may pass through a number of stages before being released. Visitors may arrive through ambulance or the main reception room, for example.

At this point, patients' information is recorded into the primary emergency department administration system, and they are either admitted (in life-threatening circumstances) or routed to the waiting area, depending on their condition. An experienced nurse triages the patient after a targeted waiting time of less than fifteen minutes and chooses the most appropriate course of action. The Manchester Triage Scale, which is utilised in all Northern Ireland hospitals, is used to identify which patients are given the highest priority. Identifying which patients need urgent medical treatment and which patients may be left to wait comfortably [18] is an important part of the process. In the patient journey, triage is a critical stage since it allows for the most effective use of resources while also increasing patient satisfaction and safety [19]. In spite of this, triage systems have been proven to be accurate in predicting hospital admission; nevertheless, they are most accurate at either end of a scale and less accurate for the overwhelming majority of patients in the centre, according to one study [18]. After being triaged, the patient returns to the waiting room before being evaluated by a physician, who will make a recommendation for the best course of action, which may include treatment in a hospital setting, outpatient follow-up, or discharge from the facility, depending on the circumstances. A bed request is sent by the emergency department to the appropriate ward, and the patient is placed on a waiting list after admission. Congestion in the emergency department may be caused by bottlenecks or increased demand at any point throughout the course of the procedure. A continually updated database is present at every level of the hospital administration system. This allows machine learning to predict future phases of the process, such as whether or not an admission would be required. Both of these objectives are met by the use of this knowledge in this investigation. Two goals are being pursued: the first is to design a model that can reliably predict hospital admissions from emergency departments, and a second aim is to evaluate the effectiveness of common machine learning algorithms in predicting hospital admissions from emergency departments. Also shown is the model's applicability as a decision-making and performance-management tool in the real-world setting.

## II. RELATED WORK

LaMantia et al. [9] utilised a logistic regression model to predict hospital admissions and emergency department re-attendance in elderly patients based on a combination of clinical and demographic data. In terms of properly projecting admissions, they performed well, but they failed miserably when it came to accurately forecasting emergency department re-attendance. A number of factors, including age, Emergency Severity Index (ESI) triage score, heart rate, diastolic blood pressure, and main complaint, have been shown to be the most important predictors of admission [9]. (See p. 255 for further information on this.) According to Baumann and Strout [20], it has been shown that the ESI is connected with admission of patients who are beyond the age of 65 years old. Boyle and colleagues [2] built prediction models for emergency department presentations and hospitalizations based on historical data collected from a large number of patients. For the purpose of evaluating model performance, the mean absolute percentage error (MAPE) was used. The best attendance model had a MAPE of approximately 7 percent, and the best admission model had a MAPE of approximately 2 percent for monthly admissions, with the best admission model having a MAPE of approximately 7 percent. However, in contrast to relying solely on historical data to forecast future events, which has the advantage of allowing projections to be made much further in advance, it has the disadvantage of overlooking data collected at the time of admission and during triage, which could improve the accuracy of short-term admission forecasting. A logistic regression model was built by Sun et al. [8] that predicted the chance of admission at the time of triage based on two years of routinely accessible administrative data gathered during the course of their study. Age, ethnicity, arrival mode, patient acuity score, underlying chronic diseases, and past ED visits or hospitalizations during the previous three-month period all increased the likelihood of admission. However, despite the fact that their data suggested that females were accepted at a greater rate than boys, the nal model made no distinction between the two. On the basis of two years of routine administration data from Glasgow hospitals, Cameron et al. [11] created a logistic regression model that may predict the probability of admissions at triage. They then used the model to verify their hypothesis. The researchers determined that the most important variables in their model were "triage category, age, National EarlyWarning Score, ambulance arrival, referral source, and admission within the previous year" based on an area under the curve of the receiver operating characteristic (AUC-ROC) of 0.877. (pp. 1). The odds ratios were not high enough to be included in the final models despite the fact that attendance during the workday and out-of-hours, as well as female gender, were statistically significant. A logistic regression model was used by Kim et al. [21] to estimate the frequency of emergency admissions at a hospital using administrative data and a logistic regression model. They had a lower level of precision, with a best accuracy of 76 percent at its most accurate point, compared to the other group's. Contrary to what these models suggest, Xie [22] reported that a Coxian Phase model outperformed a logistic regression model, with the former achieving an AUC-ROC of 0.89 and the latter achieving 0.83, indicating that the former is better. Using fuzzy min-max neural networks (FMM) to predict admissions to the emergency department, Wang et al. [23] compared the effectiveness of fuzzy min-max neural networks (FMM) to other traditional data mining approaches such as classication and regression trees (CART), Multi Layer Perceptron (MLP), random forest, and AdaBoost. They discovered that FMM was much more successful than the other treatments tested. A genetic method was employed to correctly predict 77.97 percent of cases using MLP and Random Forest, while MLP and Random Forest were also accurate in identifying 80 percent of instances using a genetic strategy. The authors of Peck et al. [24] built three models to predict emergency department admissions, using logistic regression models, naive Bayes, and expert opinion, among other techniques. In this research, it was shown that all three techniques were effective in predicting emergency department admissions. Variables such as age, type of arrival, emergency severity index, designator, major complaint, and ED provider, among others, were taken into consideration. It was determined that their logistic regression model was the most accurate in terms of predicting emergency department admissions with an AUC-ROC of 0.887. In a surprising turn of events, our model surpassed the triage nurse's prediction of the likelihood of admission. Further research revealed that the use of logistic regression to predict admission was transferable to other hospitals [10], which was confirmed in a subsequent study. It was discovered that patients spent less time in the emergency department when Peck and colleagues [25] utilised simulation models to illustrate how applying predictive models to prioritise patient discharge or treatment may minimise the amount of time they spend there. Patients in the emergency department were classified as either being discharged or being admitted to one of three distinct hospitals, according to Qui et al. [26]. The entire quality of their model was praised, and they obtained a good rating.

The accuracy in this example was 91.9 percent, while the AUC was 0.825 percent. Alternatively, the precision with which the target ward could be predicted varied based on the ward and the probability threshold that was utilised. Using eight simple machine learning algorithms, Lucini et al. [27] were able to predict emergency department admissions based on features extracted from the text of the patient's medical record in a recent research. On average, six of the eight strategies performed similarly, with AdaBoost and a decision tree proving to be the most difficult to use. In this research, six of the eight techniques performed similarly to one another: nusupport vector machines, support vector classication, extra trees, logistic regress, random forests, and multinomial nave bayes. The other two approaches fared significantly worse than one another. In the study conducted by Cameron et al. [28], a novel approach was used in which the accuracy of nurses' emergency department admission projections was compared to the accuracy of an objective score. They are more accurate than objective scores in cases when nurses feel certain that a patient will be admitted; but, in instances where nurses are unsure of whether or not the patient will be admitted, they are less accurate than the objective score.

For the purpose of arriving at a conclusion, the study emphasises the use of a combination of classical and machine learning approaches to anticipate outcomes.

Inpatient admissions to the emergency department under a variety of circumstances, based on a number of datasets This study, on the other hand, may be able to fill in some of the gaps that currently exist in the existing studies.

The inquiry has shown to be advantageous. The bulk of previous research centred on a restricted range of approaches, principally logistic regression, with just a few studies examining additional techniques. This has changed in recent years. As a consequence, it is possible to design more accurate prediction models that are based on different algorithms as a result of this. It should be noted, however, that no gradient boosted machines (GBMs) were used in any of the research studies reviewed, despite the fact that GBMs have been shown to be useful in predicting binary outcomes in other circumstances, such as hospital transfers and mortality [29]. Furthermore, just a few studies were conducted in the United Kingdom that focused on the environment, and none were conducted in Northern Ireland that focused on emergency department visits. Considering that the form and function of health care varied substantially among countries and across regions within countries, there is a critical need for more study in this field. Although the vast majority of prior research has focused on predictive models for a single hospital site, there have been just a few studies that have integrated data from several hospitals. This programme seeks to make a significant contribution to the current body of knowledge via the construction of machine learning models based on a unique dataset and the comparison of the performance of less often used approaches to the more regularly used logistic regression methodology. It is also conceivable, given that the data employed in this study is often available at the time of triage, to construct a fully automated decision support system based on the models offered here that will be totally automated.

# III. METHODS

Seven data mining tasks were carried out on the data collected from the participants in this study. In order to make predictions on the test data, the following steps were taken: extraction, data cleansing, feature engineering, data visualisation, descriptive statistics, data splitting into training (80 percent) and test (20 percent) sets, model tuning, 5-fold cross validation, predictions on the test data, and finally model performance evaluation. They aid in the optimization of models as well as the avoidance of overttting over the course of the procedure. Administrative data was employed in the study, which was acquired via electronic systems and afterwards warehoused for use in business intelligence, analytics, and reporting applications. Data from two major acute hospitals in Northern Ireland that are part of a single health and social care trust was collected over the course of the 2015 calendar year. The data includes all emergency department visits at the two hospitals, which are both part of a single health and social care trust. The information was acquired during the course of the calendar year 2015. In addition to two big acute hospitals where the research was conducted, it maintains a number of other institutions that offer a broad spectrum of acute and community health and social care services to the general population. Besides offering an extensive range of inpatient, outpatient, and emergency treatments, both hospitals have significant linkages to other parts of the healthcare system, such as the community and social services. Hospital 1 is the larger of the two institutions, with an annual capacity of more than 60000 inpatients and day cases, as well as 75000 outpatients. A total of approximately 20,000 inpatients and day cases, as well as 50,000 outpatients, may be accommodated at Hospital 2. Located in the suburbs, Hospital 1 provides medical care.

Using the information gathered at each phase of the patient journey and stored on the main administrative computer system at the time of the event's occurrence, as stated above, this model was developed and tested. It was selected which final variables to include based on the findings of prior research, the usefulness of the model, and the impact of incorporating them on the model's performance, among other considerations. Patients' age and gender, the arrival model utilised, their care group, their Manchester triage category, and whether or not they had been admitted to a hospital in the preceding week, month, or year were all taken into consideration by each of the nal models. It is possible to belong to a variety of various sorts of care groups, and each one has its own path that a patient should take. It is a scale that measures the severity of an issue and is used to establish its priority in a specific circumstance, which is known as the Manchester triage category. It was able to discover earlier admissions objectively by running a query on a hospital database. The date of attendance was further split down into components that were significant to the year, day of the week, and month of the year, among other things, using feature engineering. Every one of these models included a dependent variable that measured hospitalisation after admission from the emergency department. The vast majority of the model's variables must be input into the ED system using drop-down menus to be effective. ED System (Emergency Department): In the end, the dataset was rather clean for analysis, thanks to the listwise deletion of duplicates that had been undertaken prior to the start of the study.

occurrences in which data is not immediately accessible In this research, patients who are sent to direct assessment units or observation units are not included since they travel a different path than patients who are transported to the main emergency room. Additionally, the fact that such departments are unavailable in so many hospitals reduces the generalizability of the findings even further Despite the fact that there were 120,600 total observations in the final dataset, 10.8 percent of them were missing data, leaving 107,545 instances for the models to be built on the remaining data. The data was divided into two datasets: a training dataset (which comprised 80 percent of the instances) and a test dataset (which contained the remaining 20 percent of the instances) (which contained 20 percent of the cases). This enabled the model to be tested and confirmed. Machine learning and exploratory study were carried out with the help of the statistical computing programme R. Data extraction and storage were accomplished via the use of a database, SQL Server (2012).

## ivMACHINE LEARNING ALGORITHMS AND PERFORMANCE

In order to construct the models, machine learning techniques such as logistic regression, decision trees, and gradient boosted machines were employed, and the models were then evaluated (GBM). In this research, logistic regression was used to provide predictions for a binary dependent variable, which might be anything from positive to negative to dead or living to admit vs. not admit, among other possibilities. For the purpose of determining the chance of a particular outcome happening in the future, the logit link function is used in conjunction with the approach. For the second round, a decision tree was used, specifically recursive partitioning from the RPART package [33], which is a component of the RPART package. Third, a decision tree was employed to organise the information. Specifically, it was Breiman and colleagues [33, 34] who made a significant contribution to the model that was used in the creation of the RPART software, which may be found at this link. Data is separated at each node according to whatever variable best separates the data. The technique is repeated until an optimal model is constructed or until the nal (terminal) nodes have a limited number of observations [33], at which point the process is done one more time [34]. The overttting of the tree may be prevented when the tree is trimmed, and the most accurate prediction model can be developed [33, 35], as previously indicated. There is a third option, known as GBM [35], which involves the creation of multiple weakly linked decision trees that are then aggregated to get the final forecast. According to several studies, the use of boosting as a forecasting strategy is more accurate than the use of a single model [35] in many situations.

The three specific algorithms that were employed to enable comparisons of regression approaches were selected specifically because they would allow comparisons of a wide variety of regularly used predictive modelling strategies, such as logistic regression, among other things,

In this inquiry, a single decision tree (RPART) and a tree-based assembly approach are used (GBM). The three algorithms that were used in this research were selected for a variety of different reasons.

With the results of this investigation, we will be able to compare the outcomes of two unique area machine algorithms (RPART and GBM) with the results of a more standard logistic regression model. If you're talking about modelling, the three approaches vary in terms of how they're carried out as well as the complexity of the nal models that are produced as a consequence of those procedures. Additionally, it was determined whether or not the method could be implemented in a real-world scenario Aspects of the dataset's characteristics had a role in the decision-making process that led to the selection of the best model. Depending on whether the circumstance is one of regression or classication, different techniques are often used, and tactics appropriate for classication were employed in this instance. After ten fold cross validation was done four times across a custom tuning grid was constructed, the model parameters associated with each approach were changed to get the desired results. This strategy is used in order to discover which of two alternatives is the most advantageous.

The use of this tool makes it easy to correct parameter values as well as prevent overtitling from occurring. [35] Resampling was used to evaluate the effectiveness of the model despite the fact that logistic regression does not allow for any modifications. Users may customise the interaction depth, the minimum number of observations in a node, the learning rate, and the number of iterations in GBM [35], while the complexity parameter and the maximum node depth are often used tuning components in recursive partitioning [36, 37, 38]. It was the CARET software that was used to train and fine-tune the machine learning algorithms, which were then utilised to make predictions from the data utilising the information obtained from the dataset. This package [35], which offers a single framework for training and updating models as well as a number of other key tools, contains a large number of useful utilities. A test dataset that had never been seen before was used to make predictions with the purpose of avoiding overtating and evaluating the models' performance. To assess the overall effectiveness of each machine learning approach, the accuracy, Cohen's Kappa, Receiver Operating Characteristics (ROC) c-statistics, sensitivity, and specicity of the method were all computed and compared. Modelling ability is measured by the AUC-ROC. Models with AUC-ROC values between 0.7 and 0.8 demonstrate acceptable discrimination ability; models with AUC-ROC values greater than 0.8 demonstrate outstanding discriminating capacity; and models with AUC-ROC values greater than 0.9 demonstrate excellent discriminating ability.

[36] Abilities

## IV. RESULTS

## A. DESCRIPTIVE STATISTICS

Table 1 of this paper contains the descriptive statistics for the dataset in question. In both hospitals, there were 24 percent of emergency department visits that led in a hospital admission, with hospital 1 accounting for 26.5 percent of visits and hospital 2 accounting for 19.81 percent of visits. Quality-wise, this is comparable to hospitals in Northern Ireland and the United Kingdom [37, 38]. Similar admission rates have been recorded in other countries, including Singapore, where 30.2 percent of emergency department visitors were admitted [8], Canada, where 17.9 percent of ED visitors were admitted [22], and the United States, where 34 percent of ED visitors were hospitalised [25]. Some of these studies, on the other hand, rely on just one hospital or a limited number of hospitals, which may not be representative of general hospitalisation trends throughout the nation. Even though the admission date was split down into multiple time periods such as days, weeks, and months in the preliminary models, the final models did not include the week of the year since it had an influence on the model's performance in the field. As a whole, weekday attendance and admissions outnumbered those received on Saturday and Sunday, with Mondays accounting for a disproportionate share of those received. According to Baker [14], attendance in the United Kingdom follows a similar pattern, with the greatest levels of attendance on Mondays and a steady fall until the end of the week. The next day, Baker [14] reports that attendance had increased somewhat throughout the weekend, with Sunday being the second biggest day of the weekend overall. People who seek treatment at the emergency room are at their lowest numbers in winter, according to the American Medical Association.

**TABLE 2.Model performance.**

| | Accuracy (%) | Kappa | AUC-ROC | Specificity | Sensitivity |
|---|---|---|---|---|---|
| **Logistic Regression** | 79.94 | 0.4600 | 0.8497 | 0.8995 | 0.5357 |
| **Decision Tree (RPART)** | 80.06 | 0.4661 | 0.8249 | 0.9015 | 0.5349 |
| **GBM** | 80.31 | 0.4724 | 0.859 | 0.9038 | 0.5379 |

In this study, a data mining approach was employed to construct and evaluate three machine learning algorithms for predicting admission robability at the triage phase of the admission process. The tactics were pitted against one another in a competition. As a whole, the GBM outperformed the other models, with the decision tree and logistic regression models lagging just slightly behind. This indicates that all three models are viable choices for deployment. Nonetheless, despite the fact that the GBM was the most accurate of the three models, the logistic regression model, because to its simplicity and ease of interpretation, may be the best option for deployment in situations where interpretability is crucial, as shown by the results of this study. This strategy, according to Kuhn and Johnson [35], is based on their suggestions and should be followed. To be more specific, they propose three approaches to identifying a model that can be implemented: Develop simpler models using more interpretable methodologies, and then consider utilising the simpler model for implementation if its accuracy is sufficient when compared to the accuracy of the more complicated model. 1. Using advanced and less interpretable models, construct the theoretically most accurate model possible. 2. In this experiment, simpler models (logistic regression and decision tree) outperformed the more complicated generalised boosting model (GBM). Consequently, the logistic regression model is straightforward to comprehend and analyse, and it clearly illustrates how many various components contribute to the projection, which may assist in garnering clinician buy-in and confidence in the forecast. Remember that while reading decision trees, it is vital to keep in mind that they are fundamentally unstable, with even little changes in data having the ability to produce large structural changes in the tree [41].

The complexity of GBMs, which merge a huge number of separate decision trees to get the final predictions, may make them difficult to understand.TheGBM, on the other hand, would be the most appropriate choice for deployment in situations where accuracy is crucial. According to past research in the literature, the models provided in this study have a higher degree of accuracy than the models published in previous research. In their work, LaMantia et al. [9] used logistic regression to model data from hospital administrative systems on patients above the age of 75, and they were able to get an AUC-ROC of 0.73 by doing so. It is their contention that their approach is insufficiently exact to enable for unique admission choices to be made in each case. Based on logistic regression, Sun et al. [8] obtained an AUC-ROC of 0.849, which is the same as that obtained by the models given here in terms of accuracy. Although they include data on pre-existing diseases, Sun et al. [8] do not achieve more accuracy than the models shown here. [Note 1] Patients with diabetes, hypertension, and dyslipidemia were shown to be more likely than the general population to be admitted to the hospital.

According to Cameron et al. [11], a logistic regression model was utilised to achieve a marginally higher accuracy, resulting in an AUC-ROC of 0.8774 in the process. It was not possible to include both of these criteria in this research because the national early warning score (NEWS), which is not used in Northern Ireland, and the referral source, which is not regularly recorded at Northern Ireland's outpatient clinic, were not available. They also covered a bigger geographic area, which resulted in a larger sample size, which may have resulted in a higher accuracy of the model overall. Through the use of descriptive statistics and the logistic regression model, it is possible to uncover a number of intriguing data patterns. In addition to the patient's age and way of arrival, triage category, care group, previous admissions, and the hospital, as well as, to a lesser extent, temporal considerations, admissions are influenced by the patient's gender and race, as well as other demographic and clinical characteristics. However, despite the fact that the data suggests that patients in more severe triage categories are more likely to be admitted, descriptive statistics indicate that patients in less severe triage categories are equally as likely as those in more severe triage categories are to be admitted. The possibility of admissions in all fields is present. Patients' situations may deteriorate after they have been triaged, or further information about their health may become available, resulting in their admission after they have been admitted. According to the logistic regression model, patients who are transported to the hospital by ambulance are more likely than other patients to be admitted to the facility. This

According to some theories, this is related to the fact that persons who are suffering from life-threatening illnesses are more inclined to call for an ambulance. This is an example of how to compare and contrast two things. A favourable relationship has been shown, according to prior studies [8, 11], between ambulance arrival and admission to a medical facility. In a similar vein, the patient's care group and triage category are likely to be proxy measures of the severity of his or her illness, according to the data. Various medical diseases may cause patients to appear at different emergency rooms at different times, which may explain the relevance of temporal and geographic differences. Rather than obtaining inference, the ultimate purpose of the study was to design predictive models. As a consequence, the correlations between the variables are noteworthy and significant in understanding the model creation process. To determine whether or whether there are any underlying reasons, more research must be conducted.

## Vii CONCLUSION

A major emphasis of this study was the development and evaluation of three machine learning models aimed at predicting emergency department admissions to hospitals. Three different data mining approaches, including logistic regression, decision trees, and gradient boosted machines, were utilised to train each model, with the data from the emergency department being collected on a regular basis for each method. Compared to logistic regression and decision trees, the GBM outperformed both of them in terms of overall performance, however the choice tree and logistic regression still performed well.

The three models presented in this study, when compared to other models published in other research, give comparable, and in some cases superior, performance than the models reported in other research. Hospital decision makers may benefit from the use of models as a decision support tool to better plan and manage resources in response to the expected patient in ow from the emergency department, as shown in this study. Patients' satisfaction may be improved as a result of this, and ED congestion may be reduced as a result, mitigating the negative repercussions of ED overcrowding while simultaneously boosting patient satisfaction.

The models have the potential to be utilised in performance monitoring and auditing procedures as well, since the comparison of projected admissions to actual admissions may be used to identify trends. Despite the fact that the model may be used to assist in planning and decision-making, decisions concerning individual admissions must still be made by a physician using clinical judgement in certain cases.

## REFERENCES

**[1] J. S. Olshaker and N. K. Rathlev, ``Emergency department overcrowding and ambulance diversion: The impact and potential solutions of extended boarding of admitted patients in the emergency department,'' *J. Emerg. Med.*, vol. 30, pp. 351_356, Apr. 2006,
doi: 10.1016/j.jemermed.2005.05.023.**

**[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] S. L. Bernstein *et al.*, ``The effect of emergency department crowding on clinically oriented outcomes,'' *Acad. Emerg. Med.*, vol. 16, no. 1, pp. 1_10, 2009, doi: 10.1111/j.1553-2712.2008.00295.x.**

**[4] 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.**

**[5] 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.**

**[6] 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.**

**[7] 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.**

**[8] 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.**

**[9] 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.**

**[10] 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. VOLUME 6, 2018 10467
B. Graham *et al.*: Using Data Mining to Predict Hospital Admissions From ED**

**[11] 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.**

**[12] 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.**

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

**[14] C. Baker. (2015). *Accident and Emergency Statistics, England*.[Online]. Available: www.parliament.uk/brie_ng-papers/sn06964.pdf**

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