ORIGINAL ARTICLE

DEVELOPMENT OF A MACHINE-LEARNING ALGORITHM FOR PREDICTING THE ARRIVAL TIME AND ASSESSING FACTORS INFLUENCING PATIENT ARRIVAL AT HOSPITAL EMERGENCY DEPARTMENT – PRELIMINARY STUDY

Ilona KARPIEL[1], Olga ZIĘBA[1]

[1] Łukasiewicz Research Network – Krakow Institute of Technology, The Centre for Biomedical Engineering, Krakow, Poland

Source of support: Own sources

Author's address: I. Karpiel, Krakow Institute of Technology, The Centre for Biomedical Engineering, Zakopianska Street 73, 30-418 Krakow, Poland, e-mail: ilona.karpiel@kit.lukasiewicz.gov.pl

Abstract: Healthcare is one of the most important topics affecting society. Emergency Departments (ED) play an important role in the patient treatment cycle, providing immediate and primary health care as well as access to services offered by hospitals. From a hospital's perspective, it is crucial that emergency departments are well organized in each hospital sector. One of the most important yet overlooked problems in the medical industry is emergency department congestion.

An attempt was made to estimate and predict, using predictive models and machine learning algorithm techniques, the impact of factors on the arrival time of patients at hospital emergency department and to verify the effectiveness of teaching the models, along with a comparison which one is the best in predicting the phenomena. Research materials were source data from the 2018 National Hospital Ambulatory Medical Care Survey (NHAMCS) – Emergency Department.

In this paper, a study was conducted to search for factors that could affect the frequent arrival of patients at emergency departments. A tool for predicting the expected arrival time, taking into account the current arduous conditions/factors in the ED, has the potential to improve situational awareness and contribute to mitigating problems related to congestion. It appears that the model can be used to enhance decision support systems by determining patient arrival times.

The results obtained in the random forest regression algorithm are estimated at R2 of 86% and RMSE of 5.45. For research purposes, when attempting to analyze on a sample and reliable hospital department database, the random forest regressor method identified the most relevant factors such as initial vital signs.

The results obtained allow for a broader view in the context of assessing the prevailing situation in the emergency department. In this case – examining the most relevant factors influencing patient arrival times. With the results of high effectiveness, an algorithm can be designed to assist emergency departments in proper monitoring of existing problems.

Keywords: machine learning, databases, hospital system, data analysis, Emergency Departments

INTRODUCTION

Healthcare is one of the most important topics affecting society. Emergency departments play an important role in the patient treatment cycle, providing immediate and primary health care as well as access to services offered by hospitals [4,47]. Taking care of patients in an appropriate and timely manner, as well as making the right decisions regarding patient admission, is a major challenge for healthcare services [4]. From a hospital's perspective, it is crucial that emergency departments are well organized in each hospital sector. Given its key position in the organizational structure of the hospital, a poorly functioning Hospital Emergency Department (HED) affects accurate decision-making in the hospital unit [3]. One of the most important yet overlooked problems in the medical industry is emergency department congestion [16,18,29]. Emergency departments in European countries and the US are seeing record numbers of patients presenting [47]. These are the most severely injured individuals or those requiring immediate care. It is often very difficult to immediately determine the condition of all patients present in the emergency department when rapid intervention is required. Actions taken under the pressure of increasing life-threatening risks and the number of patients waiting for help result in wrong decisions and sometimes delays that quickly lead to hospital overcrowding [37]. When this occurs, hospitals are unable to admit patients on time, or the waiting time for treatment increases significantly [5]. The problem of hospital overcrowding may seem easy, but is is actually very difficult to solve [37]. Patients who have been admitted but have to wait a long time in the emergency department, often with high anxiety levels, can lose trust in healthcare systems. If emergency departments function poorly, this threatens not only the health and safety of the patient, but also the public's trust in the health service [3]. Another important aspect raised in the effective functioning of the emergency department and efficient recording of patient flow is the triage. A proper triage system that categorizes patients into correct priorities is equally key to increasing safety and better management of emergency patients. The current triage system in Poland is not sufficiently effective [35,47]. It is therefore necessary to develop an innovative approach to solve this global problem so as to improve patient flow and prevent the reorganization of hospital operations [37]. The health service must provide care to a large number of patients, many of whom are in critical condition. Healthcare professionals should be able to quickly access patient information and clinical data for immediate decision-making [47]. However, a person's ability to multitask is very limited and making a diagnosis is often difficult. An effective and robust methodology would allow early detection of diseases and could be used by physicians as a decision aid. Therefore, disciplines such as statistics and computer science are essential in supporting medical research. These disciplines are faced with the challenge of discovering new techniques beyond the traditional ones [4]. An essential tool supporting physicians and ED staff is artificial intelligence [47], especially machine learning (ML) [9,13,21,22,40,43,44], which utilizes various algorithms. The past five years have seen an increase in the number of applications of Artificial Intelligence (AI) and machine learning in various sectors of the economy, including healthcare, which has yielded many impressive advances, from autonomous driving to drug discovery. With the development of AI, algorithms, and machine learning technologies that are essential in searching for solutions to problems in the medical sector, various opportunities have emerged to guide future development efforts. Just a few years ago, despite enthusiasm among researchers, acceptance of new technologies in the medical space was both enthusiastic and cautious. The introduction of new technologies into a complex system in the medical sector often yields unpredictable results. Despite this uncertainty, it is optimistic to look at new technologies through the lens of existing treatment paradigms to predict how patient outcomes can be affected and potentially improved. The current generation of machine learning systems in medicine is largely aimed at limited diagnostic aids in radiology, cardiology and pathology. Examples of AI-based systems include Computer Tomography (CT) brain hemorrhage detection, systems for detecting lung nodules and coronary calcification, as well as echocardiography tools. Existing advanced systems tend to focus on problems for which a solution is readily achievable within the timeframe of the treatment cycle. This has led to the development of systems with varying degrees of clinical utility in emergency departments. It can be said that the direction of the thriving implementation of artificial intelligence in emergency departments is the same as in other medical and non-medical fields. The initiation of AI deployments will provide access to systems that are clinically relevant but narrowly applicable. As systems mature, some routine patient care tasks will be automated and decision-making in cognitive tasks will be aided by AI systems [31].

The research problem is an attempt to predict and estimate, using the development of predictive models and machine learning algorithm techniques, the impact of factors on the time of patients arrival at the hospital emergency department, and to verify the effectiveness of teaching the models, along with a comparison which one is the best in predicting the phenomena.

MATERIALS AND METHODS

Source data from the 2018 National Hospital Ambulatory Medical Care Survey (NHAMCS) – Emergency Department, which was made publicly available on the CDC government website (www.cdc.gov) as an IBM SPSS data file, was used as research material. The NHAMCS survey is based on outpatient visits to physicians and hospital emergency rooms over a 12-month period and is designed to collect data on the use of outpatient care services. It covers all fifty U.S. states and the District of Columbia and is a nationwide sample of non-institutional general and short-stay hospitals, excluding federal, military and veterans hospitals. The original database contains a total of 20,291 records and 946 columns. The overall data analysis, including data preparation and the development of predictive models, was performed using the Jupyter and RStudio environments, which use Python and R languages with libraries for analysis and design of machine learning algorithms. The study applied machine learning and used algorithms for supervised learning to solve the set objective. Regression algorithms were used to develop predictive models. The data collected were sorted and assigned to the appropriate groups, as shown in Figure 1:

A	AI	AJ	AK	AL
Variable	IMMEDR	PAINSCALE	SEEN72	RFV1
Question	Immediacy with which patient should be seen (unimputed)	Pain scale (0-10)	Was patient seen in this ED within the last 72 hours?	Patient's complaint, symptom, or other reason for visit
Subquestion				
Scale	Unknown	Unknown	Unknown	Unknown
Missings				
Values/Labels	-9 | Blank	-9 | Blank	-9 | Blank	-9 | Blank
	-8 | Unknown	-8 | Unknown	-8 | Unknown	10050 | Chills
	0 | No triage for this visit but ESA does conduct triage		-7 | Not applicable	10100 | Fever
	1 | Immediate		1 | Yes	10120 | Other symptoms of body temperature
	2 | Emergent		2 | No	10121 | Feeling cold
	3 | Urgent			10122 | Feeling hot
	4 | Semi-urgent			10123 | Feeling hot and cold
	5 | Nonurgent			10150 | Tiredness, exhaustion
	7 | Visit occurred in ESA that does not conduct nursing triage			10200 | General weakness
				10250 | General ill feeling
				10300 | Fainting (syncope)
				10350 | Symptoms of fluid abnormalities
				10351 | Edema
				10352 | Excessive sweating, perspiration
				10353 | Excessive thirst
				10400 | Weight gain
				10450 | Weight loss
				10451 | Recent weight loss
				10452 | Underweight
				10460 | Symptoms of face, not elsewhere class...
				10500 | Chest pain and related symptoms
				10501 | Chest pain
				10502 | Chest discomfort, pressure, tightness
				10503 | Burning sensation in the chest
				10550 | Pain, specified site not referable to...

Rawdata | ListVariables | Source

Fig. 1. Excerpt of coding in the data file (own elaboration). Screenshot from own research.

Data Preparation

The downloaded SPSS .sav database source file was first converted to a Microsoft Excel spreadsheet file. The database, an excerpt of which is shown in Figure 2, comprises 948 columns, which were then truncated to 203 columns and 19,351 rows since the removed portion represented empty records. Features that are redundant and will not be relevant to the analysis were omitted. The spreadsheet thus reworked serves as the basis for working on the data for further analysis. The data preparation and analysis process for further predictive analysis involved loading the necessary libraries for specific functions required in further analysis, including analytical libraries, feature selection functions, individual types of predictive models, and predictive statistics of the models. In addition, the subsequent steps checked whether the database had actually been coded correctly. The database contained negative values, which according to the coding meant empty. Negative values were converted to empty for the sake of preserving the principle of feature transformation, and then the data cleaning process was carried out. This stage of data mining is very important because in the further modeling process, machine learning algorithms underperform or completely fail if the input data, through its variability, introduces unwanted distortions and noise. The arrival time target variable, which was encoded in military time format, was then verified. Such a format is impractical for analysis, so an hourly record was extracted. This format is more suitable for analysis, yielding more reliable and standardized results. The correctly prepared file formed the basis on which the target analysis was carried out.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19351 entries, 0 to 19350
Columns: 203 entries, arrival_time_nums to patient_code
dtypes: float64(64), int64(139)
memory usage: 30.0 MB
```

	arrival_time_nums	month_of_visit	day_of_week_of_visit	wait_time_to_first_provider	length_of_visit_in_minutes	patient_age_in_years	age_recode	age_
0	20	12	1	21.0	93.0	5	1	
1	19	12	1	12.0	48.0	5	1	
2	14	12	6	21.0	99.0	0	1	
3	17	12	4	59.0	493.0	21	2	
4	17	11	2	25.0	117.0	26	3	

5 rows × 203 columns

Fig. 2. Database (own elaboration). Screenshot from own research.

Selection of Relevant Features

After the exploratory process, we proceeded to generate a function that would return the most important features in the clustering method in order to determine which of them has the greatest impact on multidimensional scaling. Correlations of the independent variables were then verified. Two methods were chosen to select the most relevant variables and the best data sample. The first method is correlation, i.e. calculating the significance of the effect between the variables x and y. The correlation results show that the data varies too much, which means that common features cannot be easily identified in the analyzed database. The second method, namely PCA, is a feature clustering technique. It is one of the so-called sensitivity analyses and involves demonstrating significance based on variance. The database was clustered into 12 divisions. Next, the cluster that provided the best results in terms of the variance of relationships in the clustering of these variables was calculated. The variance of the variables is referred to as variance. The most significant variables with the highest variance were selected for further analysis. Figure 3 below shows a set of the most relevant variables from the entire range of features in the database, which will be used in further data modeling analysis.

	PCA_Value	Variable
0	62185.230416	medication_1
1	45055.997491	patient's_complaint_symptom_or_other_reason_fo...
2	42913.881816	patient's_complaint_symptom_or_other_reason_fo...
3	1340.348556	index
4	20.471795	length_of_visit_in_minutes
5	19.619153	length_of_stay_in_observation_unit_in_minutes
6	15.843903	hospital_number
7	8.339000	initial_vital_sign_heart_rate_per_minute
8	6.262934	wait_time_to_first_provider
9	4.668952	initial_vital_sign_temperature

Fig. 3. The most relevant values (own elaboration). Screenshot from own research.

Breakdown of the Data Set

Initially, the explanatory variables, or model predictors x and y, were defined. Then, in order to be able to teach and test the effectiveness of generating predictors, the dataset was randomly divided into a training dataset and a test dataset, in which 30% of the data was taken to test the model. The training set consisted of the remainder of the data used to build the model. The following is an excerpt from the code (Figure 4) for the breakdown of the set used in the predictive model.

```
y = databank.czas_przybycia_nums
X = df_pca
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)
```

Fig. 4. An excerpt from the code of the algorithm used to break down the set. Screenshot from own research.

Training and Model Selection

Regression models are among the most widely used, practiced and developed methods in the context of analyzing large data sets. The main idea of regression is the process of predicting and forecasting data for a particular variable in relation to other variables. In other words, it is the process of fitting a variable to a newly created one according to a previously learned pattern. Prediction is possible once a regression model has been built on the basis of analysis, which will predict the value of a given characteristic based on an assumed statistical error [36]. A description of the regression models that were used in the analysis is presented below:

Linear regression – a basic type of regression in statistical analysis. It assumes a linear relationship between the explained variable and the explanatory variable, similar to correlation. It is assumed that an increase in one variable causes an increase or decrease in the other. The regression function in this case is linear [34].

Ridge regression – another approach to estimate the coefficients of models with highly correlated independent variables. It was developed to address the issue of imprecise least squares estimators in linear regression models where the independent variables are highly correlated with other independent variables. Therefore, as a solution, a ridge regression estimator was created to provide a more accurate estimation of the ridge parameters. Variance and mean squared estimators are often smaller than the previously obtained least squares estimators [46].

Decision tree – builds regression or classification models in the form of a tree structure. It breaks the dataset into smaller and smaller subsets, while incrementally expanding the associated decision tree. The end result is a tree with decision nodes and leaf nodes. A decision node has two or more branches, each representing values for the attribute under study. A leaf node represents a decision for a numerical target. The highest decision node in the tree, which corresponds to the best predictor, is called the root node. Decision trees can handle both categorical and numerical data [48].

Random forest regression – each decision tree has a high variance. However, if each of these trees is combined together in parallel, then the resulting variance is low because each decision tree is perfectly trained on that particular sample of data and therefore the result does not depend on one decision tree, but on many. For a classification problem, the final result is obtained by using a classifier with majority voting. In the case of a regression problem, the final result is the average of all results. Random forest is an ensemble technique capable of performing both regression and classification tasks using multiple decision trees and a technique called bootstrap and aggregation, commonly known as bagging or containerization. The basic idea behind this technique is to combine multiple decision trees to determine the final outcome, instead of relying on individual decision trees. A random forest has multiple decision trees as underlying learning models. Random sampling of rows and features from the dataset is performed, creating sample datasets for each model [49].

Gradient enhancement – enhancement algorithms are among the most common and widely used algorithms. They are considered one of the most powerful predictive modeling techniques. The basic principle of boosting, like other clustering algorithms, is to combine several weak "students" into a single, stronger one. It involves using basic machine learning models, sequentially one after the other, that fail to adequately predict outcomes. Each subsequent model attempts to correct the errors of its predecessor. Eventually, the models are combined to build one strong model [23].

Research scenarios

In order to extend the analysis regarding the prediction of visits or emergency arrivals to hospital emergency departments, calendar variables such as time of day, day of the week, month or holidays and other special events, as well as mass events, can be used as predictors. In addition, variables from the environment, such as weather conditions or seasons, can be used. It can be speculated that weather conditions such as high winds, thunderstorms, tornadoes, floods and the day after rainfall may affect the ED arrival rate. An increase in the number of patients in emergency departments can also be seen on a certain day of the week or a certain time of the year. Using weekday trends and weather forecasts, emergency departments would be able to anticipate the number of patients and adjust their staff and resources accordingly. In article [6], patterns of emergency care use during different seasons was analyzed and the percentage of days with each weather factor and the percentage of visits were compared. The data indicated that the season has a strong influence on the use of medical care services in EDs, as it affects the incidence of illness and injury. In contrast, extremely cold and stormy conditions significantly reduced emergency department use, but an estimated 80–95% of expected visits occur on days with poor weather conditions [10].

RESULTS AND DISCUSSION

In this study, the predictions were based on ML models whose performance is equivalent or better than that of other studies. A study was conducted to search for factors that could influence the frequent arrival of patients to emergency departments. A tool for predicting the expected arrival time, taking into account the current arduous conditions/factors in the ED, has the potential to improve situational awareness and contribute to mitigating problems related to congestion. The model can be used to enhance decision support systems by determining patient arrival time.

The presented approach can be used as a tool to support the process of medical staff scheduling in order to effectively manage a hospital department. The proposed model can be used as an objective tool for the hospital to allocate and activate resources, e.g. preparing additional beds, calling in nurses.

Results of Quantitative and Qualitative Analysis

Based on the summary results of the predictive accuracy indicators presented in Table 1, it is possible to infer which model is best suited for determining predictions for the analyzed data set and whether it is possible to estimate by how much the model will be wrong in the next observation. The dependent variable is the arrival time, in hours. In contrast, the root means squared error of the forecast (RMSE) is given in the same unit as the target.

Tab. 1. Performance indicators of models. Own research.

Method	**R^2**	**RMSE [h]**
Linear regression	0.01517	5.53
Ridge regression	0.01328	5.49
Decision tree regression	1.0	5.49
Random forest regression	0.85838	5.45
Gradient boosting Regression	0.15707	5.45

The above results of the model accuracy assessment (Tab.1) show that the best model for predicting phenomena is the random forest regression model. This is mainly indicated by the coefficient of determination R2, the optimal score for which, according to the literature review, is estimated to be between 0 and 1, while the higher the score, the more accurate the model is and the smaller the difference between the estimated and actual value. The random forest regression model score was estimated to be 86% and the RMSE statistical error was 5.45. The results of the RMSE coefficient were selected using the GridSearchCV function, which, through a grid search method, compares all combinations of parameters and returns those with the best fit to the model. Optimizing models by tuning them with hyperparameters is one of the most important steps in machine learning. It improves model performance by finding optimal values for hyperparameters. The GridSearchCV function used is a technique for finding the best parameter values from a given set of parameter grids [2,32]. Basically, it is a k-fold cross-validation method where the given dataset is split into k consecutive folds. The set is still divided into training and testing. Each iteration keeps one partition for testing and the remaining k-1 partitions for training the model. This is repeated until each iteration has been used for testing. At each iteration, the function records the performance of the model, and finally provides an average of all the performances. This is also a time-consuming process. It is a technique for improving and enhancing model performance. The lower the score obtained, the better the model is. Figure 5 presents the result of the algorithm. The random forest regression model predicted the most relevant factors affecting the predicted target variable, such as initial vital signs-temperature and time to first medical contact.

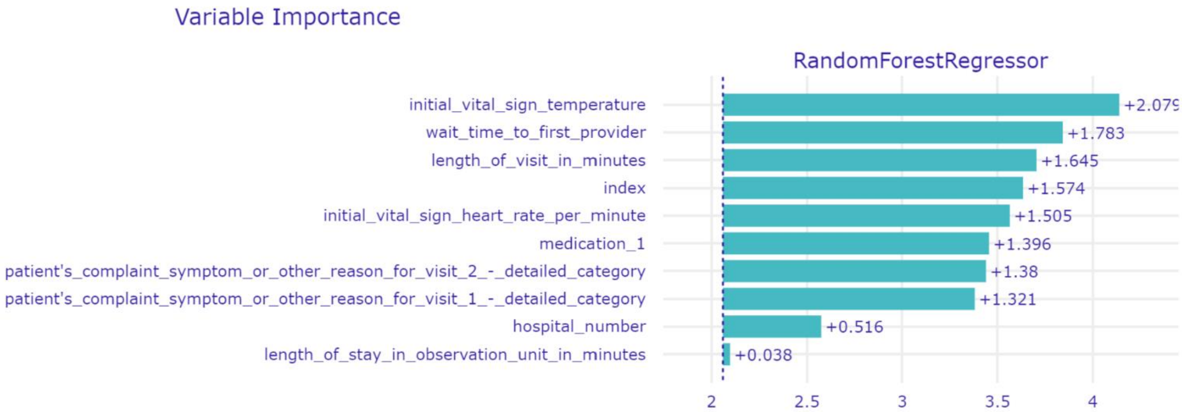

Fig. 5. A selection of the most significant factors affecting the arrival time of patients to the emergency department (own elaboration). Screenshot from own research.

Applications

The purpose of this study was to predict and estimate the impact of different factors on the arrival time at HEM through the use of predictive models and machine learning algorithm techniques. The work also had a subordinate purpose of analyzing medical data using machine learning algorithms to optimize the future management process in hospital emergency departments. The reference point for the analysis is the presentation of theoretical issues regarding the current organization of an ED, which are the cause of the malfunctioning of this department and the growth of management problems, i.e. the increase in patient congestion in the hospital departments and the lack of an organized patient flow, as well as the mishandling of operations from an organizational perspective. With an understanding of the current disorganization of the departments, possible solutions to this problem are also described, citing those already in use [5,8]. Tools used in the field of Artificial Intelligence and Machine Learning are an indispensable support in the search for solutions to problems in the medical sector. Therefore, in this paper, a study was conducted, using a reliable database, to look for factors that can influence the frequent arrival of patients to emergency departments. A tool for predicting relevant factors affecting the arrival time would have the potential to improve situational awareness and contribute to counteracting crowding problems. In order to search for the best model, a comparative analysis of five Machine Learning (ML) algorithms (linear regression, ridge regression, decision tree regression, random forest regression, gradient enhancement) that differ in their training mechanism was conducted. As a result of the analyses, the random forest regression model was found to be the best-fitting model for predicting significant factors affecting patient arrival time with the best coefficient of determination score of 86% and the lowest RMSE error rate of 5.45. It was noted that only one model was the best in terms of predictive performance, training time and interpretability. The random forest method identified the most significant variables such as initial vital signs-temperature and then time to first medical contact. These variables appear to be of little importance to the problem posed when determining the impact of individual factors on arrival times, while they performed well in training the model, as confirmed by the results of the predictive statistics obtained. The other models did not prove their effectiveness. Due to an overly diverse database in terms of divergence and randomness of individual patient values, and the lack of additional variables such as weather conditions, which could highlight new patterns unattainable by the tested model, it is hard to pinpoint the time series in which one might expect a higher intensity of ED visits. In summary, the results indicate that ML algorithms can accurately predict the impact of factors on patient arrival times to HEDs. The analysis of the selection of relevant features by applying ML algorithms for patient arrival time not only fulfilled one of the research objectives but was also used in further analysis – Machine Learning. The results estimated by the algorithms indicate a correctly selected predictive modeling technique, which worked well with the variables contained in the database. The solution used in the paper was presented as an example of the use of ML techniques aimed at using the solution in emergency departments. The ML approach presented herein can serve as a tool to assist in building medical staff schedules and can be integrated, for example, into a human resources system to effectively manage and mitigate the utilization of hospital emergency department resources. ML models can also be applied to, for example, symptom-targeted data, enabling earlier interventions using advanced diagnostics and tailoring better and cost-effective personalized therapies. The proposed model can be used as an objective tool for the hospital to allocate and activate resources, e.g. calling in nurses and doctors, preparing additional beds for patients, and consulting with admitting physicians about alternative treatments in other hospital units. In practice, the forecasting model can be put into everyday clinical use by augmenting existing decision support systems that track patients in the ED or by adding an estimate of the number of patients expected to arrive within a certain time frame.

Measures to improve the efficient management of HEDs play an important role in improving the quality of health care. Of particular importance is the knowledge of what solutions can be introduced, and what solutions applied in other countries, for example, have proved successful. The phenomenon of overcrowding in emergency departments, which is common both in Poland and worldwide, has led to a desire to solve this problem. The research work sought a pattern through the use of Machine Learning techniques.

Arrival time to the HED

Arrival time is of great importance to healthcare [8,42,50,51]. Patient flow through emergency departments (ED), overcrowding and long waiting times are acknowledged as increasing worldwide issues in healthcare [11,52]. Overcrowding in emergency departments is a serious problem in many countries. Accurate ED patient arrival forecasts can serve as a management baseline to better allocate ED personnel and medical resources. Zhang et al. [45] presented two methods of features for forecasting patient arrivals. Their results showed that for hourly forecasting of patient arrivals, each machine learning model performs better than the traditional Auto Regressive Integrated Moving Average (ARIMA) model. Accurate information about arrival times, the number of patients that can reach the hospital, can be used as a basis for management to better allocate staff, adapt rooms or free up beds [1,12,20]. Patient arrivals are determined by many variables and are extremely unpredictable. However, research shows that machine learning algorithms can significantly improve time prediction performance [19,25,28,38] and a number of proposed solutions is growing [7,14,24,26,41].

The arrival time issue is also analyzed in relation to patients with acute ischemic stroke. There is a great need to find a way to reduce the time to get to the hospital and to predict when patients will arrive [17]. The arrival time or the mentioned weather conditions, for example, are considered to be deeply related to the issue of patient forecasting for proactive bed management [20]. In 2022, Huang et al. [15] presented analyzes of clinical and sociodemographic factors in relation to arrival time on Kinmen Island. This study investigated factors associated with delayed emergency treatment of patients with Acute Myocardial Infarction (AMI) on Kinmen Island and offered suggestions for developing interventions to reduce the time from symptom onset to receipt of appropriate medical care [27,30].

The causes of overcrowding are important for many reasons and have implications for patient arrival times or patient handling times by medical staff [33,39,42]. It is worth noting that changes in meteorological factors beyond a certain range can cause thermal imbalances in the body, which can promote the development of many diseases [53], also affecting both the arrival time and the number of patients in hospitals.

CONCLUSION

The research objective was achieved through the use of an accurate random forest regression algorithm, as confirmed by the obtained metrics of R2 accuracy of 86% and RMSE of 5.45. For research purposes, when attempting to analyze on a sample and reliable hospital ward database, the random forest regressor method identified the most relevant factors such as initial vital signs – temperature and then the time of waiting for the first contact with a physician. To improve forward-looking analysis for predicting visits or emergency arrivals to hospital emergency departments, calendar data such as hour, day of the week or month, can be used as predictors. Also, other special events or atmospheric variables, such as weather conditions or season, can be taken into account.

The presented approach can be used as a tool supporting the process of building a work schedule for medical staff in order to effectively manage a hospital ward. The proposed model can be used as an objective tool for the hospital to allocate and activate resources, e.g. opening additional beds, and calling nurses. Our research is basic, giving some insight into the situations that it is advisable to continue.

ACKNOWLEDGMENTS

I would like to thank dr inż. Karol Jędrasiak for the helpful remarks and directions.

AUTHORS' DECLARATION

Conceptualization: Olga Zięba. **Methodology:** Olga Zięba. **Validation:** Ilona Karpiel, Olga Zięba. **Resources:** Ilona Karpiel. **Data curation:** Olga Zięba. **Writing-original draft preparation:** Ilona Karpiel, Olga Zięba. **Writing-review and editing:** Ilona Karpiel, Olga Zięba. **Supervision:** Olga Zięba, Ilona Karpiel. **Project administration:** Olga Zięba. **Funding acquisition:** Olga Zięba. The Authors declare that there is no conflict of interest.

REFERENCES

1. Afilal M, Yalaoui F, Dugardin F, Amodeo L, Laplanche D, Blua P. Forecasting the Emergency Department Patients Flow. J Med Syst. 2016; 40(7): 175.

2. Alhakeem ZM, Jebur YM, Henedy SN, Imran H, Bernardo LFA, Hussein HM. Prediction of Ecofriendly Concrete Compressive Strength Using Gradient Boosting Regression Tree Combined with GridSearchCV Hyperparameter-Optimization Techniques. Materials. 2022; 15(21): 7432.

3. Asplund K, Ehrenberg A. Triage and Flow Processes in Emergency Departments: A Systematic Review. 2010; 197: 1-34.

4. Caballé-Cervigón N, Castillo-Sequera JL, Gómez-Pulido JA, Gómez-Pulido JM, Polo-Luque ML. Machine Learning Applied to Diagnosis of Human Diseases: A Systematic Review. Applied Sciences. 2020; 10(15): 5135.

5. Casola S. Reducing waiting times and crowding in hospital emergency departments using machine learning: Master degree course in Innovation and Research in Informatics. 2018. Available from: https://webthesis.biblio.polito.it/9067/

6. Christoffel KK. Effect of season and weather on pediatric emergency department use. The American Journal of Emergency Medicine. 1985; 3(4): 327–30.

7. Coskun N, Erol R. An Optimization Model for Locating and Sizing Emergency Medical Service Stations. J Med Syst. 2010; 34(1): 43–9.

8. Derlet RW, Richards JR, Kravitz RL. Frequent Overcrowding in U.S. Emergency Departments. Acad Emergency Med. 2001; 8(2): 151–5.

9. Ehwerhemuepha L, Heyming T, Marano R, Piroutek MJ, Arrieta AC, Lee K, et al. Development and validation of an early warning tool for sepsis and decompensation in children during emergency department triage. Sci Rep. 2021; 11(1): 8578.

10. Faryar K. The Effects of W ects of Weekday, Season, F , Season, Federal Holida al Holidays, and Se ys, and Severe Weather Conditions on Emergency Department Volume in Montgomery County, Ohio. Wright State University Core Scholar; 2013.

11. Forero R, McCarthy S, Hillman K. Access block and emergency department overcrowding. Crit Care. 2011; 15(2): 216.

12. Harrou F, Dairi A, Kadri F, Sun Y. Forecasting emergency department overcrowding: A deep learning framework. Chaos, Solitons & Fractals. 2020; 139: 110247.

13. Hayashi Y, Shimada T, Hattori N, Shimazui T, Yoshida Y, Miura RE, et al. A prehospital diagnostic algorithm for strokes using machine learning: a prospective observational study. Sci Rep. 2021; 11(1): 20519.

14. Headrick RW, Morgan GW. Resource allocation in multifacility Emergency Medical Service Systems. J Med Syst. 1988; 12(3): 121–8.

15. Huang Y-H, How C-K, Ho C-S. Factors Affecting Delayed Hospital Arrival of Patients with Acute Myocardial Infarction in Kinmen. IJERPH. 2022; 19(3): 1323.

16. Iacobucci G. Overcrowding and long delays in A&E caused over 4000 deaths last year in England, analysis shows. BMJ. 2021; n2835.

17. Iosif C, Papathanasiou M, Staboulis E, Gouliamos A. Social factors influencing hospital arrival time in acute ischemic stroke patients. Neuroradiology. 2012; 54(4): 361–7.

18. Jeanmonod D, Jeanmonod R. Overcrowding in the Emergency Department and Patient Safety. In: Firstenberg MS, Stawicki SP, editors. Vignettes in Patient Safety - Volume 2 [Internet]. InTech; 2018 [cited 2022 Jul 18]. Available from: http://www.intechopen.com/books/vignettes-in-patient-safety-volume-2/overcrowding-in-the-emergency-department-and-patient-safety

19. Jiang S, Chin K-S, Tsui KL. A universal deep learning approach for modeling the flow of patients under different severities. Computer Methods and Programs in Biomedicine. 2018; 154: 191–203.

20. Jilani T, Housley G, Figueredo G, Tang P-S, Hatton J, Shaw D. Short and Long term predictions of Hospital emergency department attendances. International Journal of Medical Informatics. 2019; 129: 167–74.

21. Kim JH, Choi A, Kim MJ, Hyun H, Kim S, Chang H-J. Development of a machine-learning algorithm to predict in-hospital cardiac arrest for emergency department patients using a nationwide database. Sci Rep. 2022; 12(1): 21797.

22. King Z, Farrington J, Utley M, Kung E, Elkhodair S, Harris S, et al. Machine learning for real-time aggregated prediction of hospital admission for emergency patients. npj Digit Med. 2022; 5(1): 104.

23. Li W, Wang W, Huo W. RegBoost: a gradient boosted multivariate regression algorithm. IJCS. 2020; 4(1): 60–72.

24. Lin BY-J, Hsu C-PC, Chao M-C, Luh S-P, Hung S-W, Breen G-M. Physician and Nurse Job Climates in Hospital-Based Emergency Departments in Taiwan: Management and Implications. J Med Syst. 2008; 32(4): 269–81.

25. Liu Y, Yang C, Huang K, Gui W. Non-ferrous metals price forecasting based on variational mode decomposition and LSTM network. Knowledge-Based Systems. 2020; 188: 105006.

26. Luo L, Luo Y, You Y, Cheng Y, Shi Y, Gong R. A MIP Model for Rolling Horizon Surgery Scheduling. J Med Syst. 2016; 40(5): 127.

27. McGinn AP, Rosamond WD, Goff DC, Taylor HA, Miles JS, Chambless L. Trends in prehospital delay time and use of emergency medical services for acute myocardial infarction: Experience in 4 US communities from 1987-2000. American Heart Journal. 2005; 150(3): 392–400.

28. Menke NB, Caputo N, Fraser R, Haber J, Shields C, Menke MN. A retrospective analysis of the utility of an artificial neural network to predict ED volume. The American Journal of Emergency Medicine. 2014; 32(6): 614–7.

29. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: A systematic review of causes, consequences and solutions. Bellolio F, editor. PLoS ONE. 2018; 13(8): e0203316.

30. Moser DK, Kimble LP, Alberts MJ, Alonzo A, Croft JB, Dracup K, et al. Reducing Delay in Seeking Treatment by Patients With Acute Coronary Syndrome and Stroke: A Scientific Statement From the American Heart Association Council on Cardiovascular Nursing and Stroke Council. Circulation. 2006; 114(2): 168–82.

31. Moulik SK, Kotter N, Fishman EK. Applications of artificial intelligence in the emergency department. Emerg Radiol. 2020; 27(4): 355–8.

32. Mustaqeem Mohd, Saqib Mohd. Principal component based support vector machine (PC-SVM): a hybrid technique for software defect detection. Cluster Comput. 2021; 24(3): 2581–95.

33. Rydman RJ, Tannebaum RD, Zalenski RJ. An evaluation of Hospital Emergency Department (HED) adherence to universal precautions. J Med Syst. 1994; 18(4): 207–20.

34. Schneider A, Hommel G, Blettner M. Linear Regression Analysis. Deutsches Ärzteblatt international [Internet]. 2010 Nov 5 [cited 2023 Feb 2]; Available from: https://www.aerzteblatt.de/10.3238/arztebl.2010.0776

35. Sherafat A, Vaezi A, Vafaeenasab M, Ehrampoush M, Fallahzadeh H, Tavangar H. Responsibility-evading performance: The experiences of healthcare staff about triage in emergency departments: A qualitative study. Iranian J Nursing Midwifery Res. 2019; 24(5): 379.

36. Sperandei S. Understanding logistic regression analysis. Biochem Med. 2014; 12–8.

37. Srinivas P, Kumar DS. Prediction Of Hospital Admission Using Machine Learning. 2019; 8(12): 2764–70.

38. Sudarshan VK, Brabrand M, Range TM, Wiil UK. Performance evaluation of Emergency Department patient arrivals forecasting models by including meteorological and calendar information: A comparative study. Computers in Biology and Medicine. 2021; 135: 104541.

39. Sun BC, Hsia RY, Weiss RE, Zingmond D, Liang L-J, Han W, et al. Effect of Emergency Department Crowding on Outcomes of Admitted Patients. Annals of Emergency Medicine. 2013; 61(6): 605-611.e6.

40. Takeda M, Oami T, Hayashi Y, Shimada T, Hattori N, Tateishi K, et al. Prehospital diagnostic algorithm for acute coronary syndrome using machine learning: a prospective observational study. Sci Rep. 2022; 12(1): 14593.

41. Wullink G, Van Houdenhoven M, Hans EW, van Oostrum JM, van der Lans M, Kazemier G. Closing Emergency Operating Rooms Improves Efficiency. J Med Syst. 2007; 31(6): 543–6.

42. Yu D, Blocker RC, Sir MY, Hallbeck MS, Hellmich TR, Cohen T, et al. Intelligent Emergency Department: Validation of Sociometers to Study Workload. J Med Syst. 2016; 40(3): 53.

43. Yu JY, Xie F, Nan L, Yoon S, Ong MEH, Ng YY, et al. An external validation study of the Score for Emergency Risk Prediction (SERP), an interpretable machine learning-based triage score for the emergency department. Sci Rep. 2022; 12(1): 17466.

44. Zhai Q, Lin Z, Ge H, Liang Y, Li N, Ma Q, et al. Using machine learning tools to predict outcomes for emergency department intensive care unit patients. Sci Rep. 2020; 10(1): 20919.

45. Zhang Y, Zhang J, Tao M, Shu J, Zhu D. Forecasting patient arrivals at emergency department using calendar and meteorological information. Appl Intell. 2022; 52(10): 11232–43.

46. Zheng S. MTH 541/643: Statistical Theory II. Methods of Evaluating Estimators [Internet]. 2020 [cited 2023 Mar 24]. Available from: http://people.missouristate.edu/songfengzheng/teaching/mth541/lecture%20notes/evaluation.pdf

47. Materiały badawcze opracowane w ramach prac B+R w projekcie DOBBIO10/19/02/2020 - Opracowanie nowoczesnego modelu zarządzania pacjentem w stanie zagrożenia życia w oparciu o samouczącą się algorytmizację procesów decyzyjnych i analizę danych z procesów terapeutycznych. 2020.

48. Decision Tree Regression [Internet]. [cited 2023 Feb 2]. Available from: http://www.saedsayad.com/decision_tree_reg.htm

49. Random Forest Regression in Python - GeeksforGeeks [Internet]. [cited 2023 Feb 2]. Available from: https://www.geeksforgeeks.org/random-forest-regression-in-python/

50. Products - Data Briefs - Number 102 - August 2012 [Internet]. [cited 2023 Feb 3]. Available from: https://www.cdc.gov/nchs/products/databriefs/db102.htm

51. Financial Impact of Emergency Department Crowding - PMC [Internet]. [cited 2023 Feb 3]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3099606/

52. A ‘DURABLE OPPORTUNITY’: ED OVERCROWDING IN THE - ProQuest [Internet]. [cited 2023 Feb 3]. Available from: https://www.proquest.com/docview/1790494895?pq-origsite=gscholar&fromopenview=true

53. Application Research on Gated Recurrent Unit Deep Learning Prediction and Graded Early Warning of Emergency Department Visits Based on Meteorological Environmental Data [Internet]. [cited 2023 Feb 3]. Available from: https://www.besjournal.com/en/article/doi/10.3967/bes2020.111