Artificial Intelligence in Emergency Care: Implementing Machine Learning for Triage Optimization in Italian Hospitals*

Paulo Vitor de Campos Souza[1,*,†], Mauro Dragoni[1,†]

[1]Fondazione Bruno Kessler, Via Sommarive 18 - 38123 Trento (TN), Italy

Abstract

Accurate identification of patient emergency states in emergency rooms is vital for delivering timely and appropriate medical intervention. This paper presents a comprehensive approach using a dataset from a Northern Italian hospital to predict patient urgency levels with machine learning algorithms. We processed and analyzed the data through feature selection techniques, feature importance analysis, and interpretability methods, focusing on the dataset dimensions. Our tests resulted in accuracy exceeding 95% in three machine learning algorithms, demonstrating the feasibility of developing an intelligent computerized system capable of predicting emergency states in an emergency room setting. These findings suggest that integrating advanced data analytics can significantly enhance patient triage and hospital resource planning.

Keywords

Emergency classification, Intelligent systems, Pattern classification

1. Introduction

The identification of patient urgency levels in emergency departments (EDs) is crucial for ensuring timely and appropriate care. The Manchester Triage System (MTS) is commonly employed to categorize patients based on their level of urgency, utilizing a color-coded protocol to prioritize patient care effectively. This systematic approach aids in managing patient flow and ensuring that those who require immediate attention receive it without unnecessary delay. The implementation of the MTS has significantly contributed to the optimization of emergency department operations, enhancing patient outcomes and operational efficiency.

Recent advancements in artificial intelligence (AI), machine learning (ML), and data mining have shown promising potential in augmenting traditional triage protocols like the MTS. For example, Lee et al. [1] developed an AI model using neural networks and machine learning to predict hospital admission for urgent patients, demonstrating the potential for streamlining ED operations with minimal variables. While, Mutegeki et al. [2] utilized machine learning classification algorithms to predict the Emergency Severity Index of patients based on their medical data, demonstrating the utility of AI for enhancing triage decisions.

In our paper, we explore the application of different machine learning and data mining techniques to discover patterns in a real-world dataset of hospital visits gathered from hospitals in Northern Italy. Traditional machine learning models will be employed to analyze this dataset, which is unprecedented in its reflection of the hospital's routine in a tangible, real-world context. The contribution of this paper is twofold. Firstly, it provides a critical discourse on the dimensions impacting the screening of individuals in emergency scenarios. By applying a fusion of data science, machine learning, and data mining techniques, we unravel patterns within a genuine dataset of hospital admissions, showcasing the potential of these approaches to assist emergency room (ER) operations. Secondly, this work endeavors

*Corresponding author.
†These authors contributed equally.
pdecampossouza@fbk.eu (P. V. d. C. Souza); dragoni@fbk.eu (M. Dragoni)
0000-0002-7343-5844 (P. V. d. C. Souza); 0000-0003-0380-6571 (M. Dragoni)

to bridge the gap between theoretical models and practical applications, fostering advancements that could underpin smarter, data-driven decisions in emergency care settings.

The remainder of this paper is organized as follows. Section 2 provides a comprehensive literature review, laying the groundwork for the applied methodologies and situating our work within the existing body of research. Section 3 details the dataset employed in this study, including the preprocessing steps and preparation procedures essential for the subsequent analyses. In Section 4, we delve into the testing methodologies and discuss the results obtained from the machine learning models. Finally, Section 5 concludes the paper with a summary of our findings and an outline of potential avenues for future research.

2. Background and Related Work

In this Section, we present the main concepts about the emergency protocol in hospitals and some state-of-the-art machine learning techniques for solving problems in this context.

2.1. Emergency Triage System

The Manchester Triage System (MTS) is a pivotal methodology in EDs for categorizing patient urgency utilizing a five-level urgency scale to ensure patients receive care promptly based on the severity of their condition. The system is designed around a series of flowcharts, each corresponding to different presenting complaints, while leading triage nurses through a series of discriminators to assign an appropriate urgency level. This methodology facilitates a more organized patient flow within EDs, aiming to reduce wait times and prioritize care for those most in need [3].

The MTS's robust framework enhances its capability to extend beyond mere prioritization, offering insights into patient outcomes such as hospital admission rates and mortality. Studies have demonstrated the system's effectiveness in distinguishing between high and low-risk patients, proving its value not only as a triage tool but also as a predictive model for patient disposition [3]. Moreover, the MTS's adaptability to various healthcare settings underscores its utility in improving emergency care delivery worldwide.

To illustrate the classifications within the MTS, Table 1 summarizes the color-coded urgency levels and their meanings.

Color Code	Meaning
Red	Immediate attention required
Orange	Very urgent
White	Urgent
Green	Standard
Blue	Non-urgent

Table 1
Manchester Triage System Color Codes and Their Meanings

2.2. Machine Learning Models and Datasets in Emergency Detection Using the Manchester Triage System

A critical component of ED efficiency is the accurate triage of patients based on urgency, for which the MTS is widely used. Recent studies have integrated ML techniques to enhance the predictive power of MTS, improving patient outcomes and resource allocation.

Roquette et al. (2020) explored deep neural networks (DNNs) combined with triage textual data to predict ED admissions, emphasizing the importance of text processing for enhanced prediction accuracy [4]. Zachariasse et al. (2016) assessed the safety of MTS in pediatric care, identifying risk factors for undertriage and suggesting modifications for improvement [5]. Seiger et al. (2014) examined modifications to MTS and the inclusion of abnormal vital signs in pediatric EDs across multiple centers, aiming for reduced incorrect assignments [6]. Azeredo et al. (2015) reviewed the efficacy of MTS for risk classification, pointing out its validity across various patient demographics but also its tendency towards sub-triage [7]. Zachariasse et al. (2017) evaluated MTS's validity across different emergency care settings, noting moderate to good performance with lower effectiveness in young and elderly patients [8]. Santos et al. (2013) investigated MTS version II for resource utilization, indicating improvements, especially in surgical specialties [9].

3. Dataset: Emergency protocol in an Italian hospital

This section outlines the methodology adopted to prepare the dataset we obtained from a Northern Italian hospital and that has been used for our investigation. The dataset comprises various dimensions, including patient codes, gender, date of birth, residency, age, mode of arrival, primary problem, diagnosis codes, and the urgency of care required. The objective was to predict the state of emergency from these features. The methodology employed is described step by step below:

3.1. Data Cleaning and Preprocessing

Data cleaning and preprocessing are pivotal initial steps to ensure the quality and coherence of the dataset for subsequent analysis. This process lays the groundwork for a robust predictive model by enhancing data integrity and relevance.

The dataset, originating from the automated system of a hospital in Northern Italy, comprises 582 instances and 18 features with approximately 17% missing data. To enhance data quality and prepare for analysis, features, and samples with missing values were meticulously identified and pruned from the dataset.

Conforming to patient privacy regulations, any variables that could potentially lead to patient identification were rigorously excluded. In alignment with privacy considerations, the following identifiers were removed: Encounter ID, Fiscal Code, and Birth Date. Moreover, to establish an unbiased foundation for model comparison against human-assigned urgency levels, the Urgency Triage dimension was omitted to prevent it from influencing the model's prediction of the final emergency state. Similarly, Outcome and Send Mode, which relate to the patient's ultimate stage in the care pathway, were excluded, as their inclusion post-triage could introduce post-hoc bias into the predictive model. The Diagnose Code was also excluded, as it is typically generated after the triage process and does not reflect the initial presentation.

The remaining dimensions retained for analysis include Commune Residence, Age, Arrival Mode, and Main Problem (1) and (2). These dimensions were selected for their relevance to the patient's initial clinical presentation and their potential utility in predicting the appropriate level of emergency care.

By concentrating on these dimensions, we aim to build a predictive model that mirrors the initial assessment of a patient's emergency state based on unbiased, non-identifiable clinical features, thereby allowing a direct and fair comparison with the hospital's triage outcomes.

After the data pre-treatment procedures, 5 dimensions remained, which are shown in Figure 1.

3.2. Target Variable Definition

The core of an effective triage system is its ability to quickly and accurately determine a patient's condition urgency, guiding care prioritization. Our research utilizes data mining to identify urgency as the key target variable, crucial for the patient care process. This variable, hereafter called *Appropriateness Outcome*, stems from a thorough triage evaluation of the patient's condition severity and immediacy. It is essential for training algorithms to accurately predict case urgency, influencing critical decisions like resource allocation, tending to immediate medical needs, and managing patient flow in emergency care. Figure 2 graphically shows the relationship between age and emergency classification.

	Name	Distribution	Mean	Median	Dispersion	Min.	Max.	Missing
N	Municipality of Residence		16.1993	11	1.01615	0.00	54	0 (0%)
N	Age		47.80	53	0.57	0	97	0 (0%)
N	Arrival Mode		1.71821	2	0.283182	0.00	3	0 (0%)
N	Main Problem (1)		4.61512	2	1.14922	0.00	13	0 (0%)
N	Main Problem (2)		4.61512	2	1.14922	0.00	13	0 (0%)
C	Appropriateness Outcome			GREEN	0.948			0 (0%)

Figure 1: Feature statistics-Hospital emergency prediction dataset.

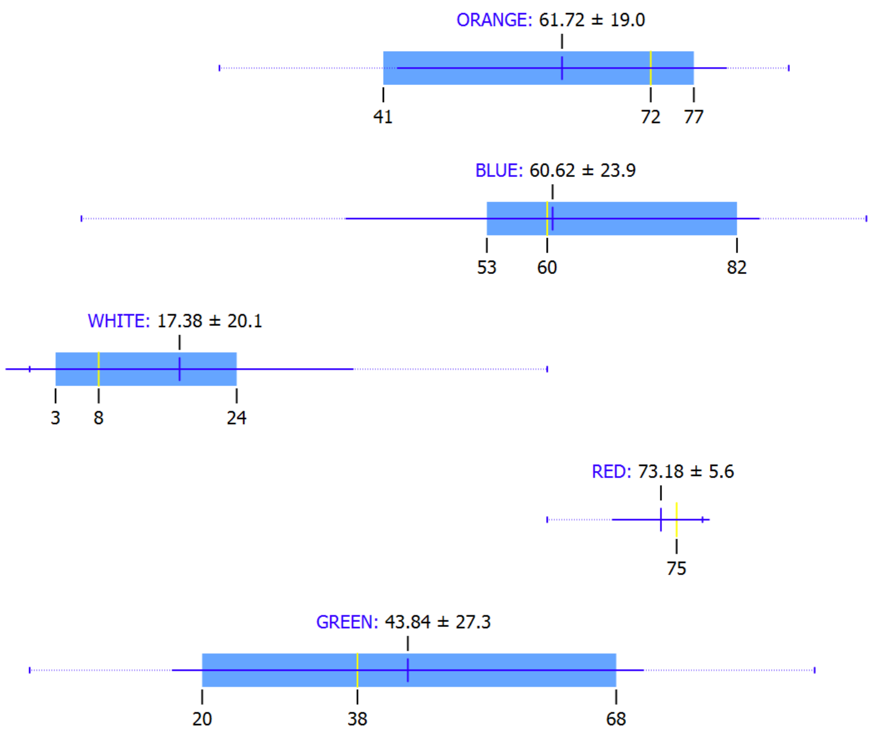

Figure 2: Feature correlation-Hospital emergency prediction dataset.

The complexity and insight of the relationship between the target variable and input features, such as presenting symptoms, age, and arrival mode, directly influence the assigned urgency level of a patient's case. For example, specific symptoms may indicate life-threatening conditions requiring immediate care, while others are less severe. Age reflects a patient's vulnerability to certain emergencies and complications, affecting urgency classification. Similarly, arrival mode suggests the severity of the condition, with ambulance arrivals often indicating a need for urgent care. This understanding, captured in Figure 3, enables future models to replicate the nuanced decision-making of human-led triage. The model's predictive accuracy depends on its ability to understand these relationships, aiming to match or exceed human judgment in triage scenarios. Figure 3 illustrates the feature importance ranking, as determined by various statistical methods for feature selection, enhancing the model's effectiveness in emergency assessments.

		#	Info. gain	Gain ratio	Gini	ANOVA	χ^2	ReliefF	FCBF
1	N Main Problem (2)		0.189	0.100	0.043	15.557	48.368	0.150	0.131
2	N Main Problem (1)		0.189	0.100	0.043	15.557	48.368	0.150	0.000
3	N Age		0.210	0.105	0.032	26.886	84.779	0.147	0.143
4	N Municipality of Residence		0.107	0.054	0.022	4.742	23.468	0.104	0.068
5	N Arrival Mode		0.144	0.149	0.039	27.094	12.680	0.055	0.000

Figure 3: Feature correlation-Hospital emergency prediction dataset.

Appropriateness Outcome works not simply as a categorical endpoint, but as a vital gauge for evaluating the efficacy and quality of emergency medical services. Through our model, we endeavor to refine this gauge, harnessing the subtleties of machine learning to improve the triage process and enhance the delivery of emergency healthcare.

3.3. Feature Selection

Feature selection is a critical step in the development of predictive models, aimed at eliminating redundant or irrelevant features that could potentially degrade model performance. By narrowing down the dataset to the most informative attributes, we ensure that the models are trained on data that most significantly influence the prediction of the emergency state.

Our process for feature selection was informed by a thorough analysis of feature correlations, as illustrated in Figure 3. From left to right, the statistical techniques employed are as follows [10]:

- Information Gain: quantifies how much a feature decreases uncertainty when predicting the target variable.
- Gain Ratio: adjusts information gain for feature bias, factoring in the number and size of branches it creates.
- Gini Index: measures a feature's ability to discriminate between classes; lower values indicate greater purity.
- ANOVA: determines if the mean of the target variable significantly differs across groups defined by a feature.
- Chi-Squared (χ^2): assesses the strength of association between features and the target variable; higher values mean greater significance.
- Relief: weights features based on their capability to differentiate between similar instances of varying classes.
- FCBF: identifies the most predictive features that are correlated with the target class while minimizing redundancy.

In Figure 3, Main Problem (2) and Main Problem (1) are evaluated similarly across most techniques, indicating their significant roles in predicting the outcome. These problems are related to the first symptom of the patient (pain, cut, fever, cold, etc). Age also shows a strong presence, especially notable in the Information Gain and Chi-Squared evaluations, which suggests its critical influence in emergencies. Municipality of Residence demonstrates moderate relevance, whereas Arrival Mode stands out particularly in the ANOVA and Chi-Squared assessments, reflecting its potential impact on the urgency classification.

Each statistical technique offers a unique perspective on feature relevance, and collectively they provide a comprehensive understanding of which features are most influential. By analyzing these rankings, we can strategically select a combination of features that will empower our predictive model to make accurate assessments reflective of real-world triage situations.

This feature selection process, therefore, is not just a step towards model optimization —it is an effort to capture the essence of practical emergency assessment, ensuring that our predictive system can operate with the highest degree of reliability and validity.

The analysis of the role of Municipality of Residence in predicting urgency in an Italian hospital context revealed methodological variations in its correlation with the Appropriateness Outcome. All dimensions were retained for comprehensive future analyses.

Figure 4 illustrates a violin plot showing how different levels of urgency, represented by specific colors, are distributed among patients from various municipalities. The width of the violin indicates the density of cases at each level of urgency, while the central point and line represent the median and interquartile ranges, respectively. This visualization aids in identifying patterns of emergency case distribution by locality, which is crucial for predicting emergency care demands and planning resources effectively. The relationship between the municipality of residence and the *Appropriateness Outcome* is essential for feature selection and maintaining the predictive accuracy of the triage system when applied across different demographic and geographic profiles.

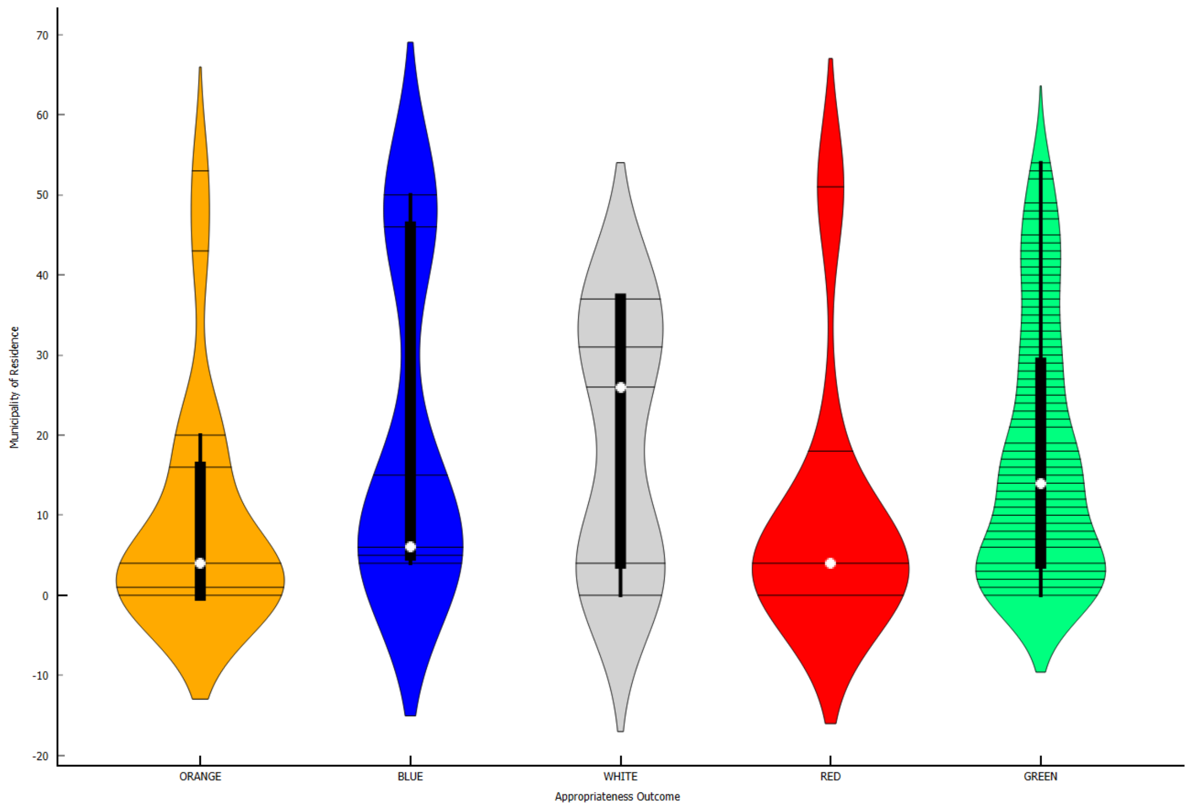

Figure 4: Violin plot from the relationship between municipality of residence and emergency classification.

By judiciously selecting features, we can refine the models' focus, directing computational power toward analyzing data points that have a substantive impact on emergency care outcomes. Such a practice not only streamlines the modeling process but also contributes to the generalizability of the model, ensuring that its applications extend beyond the immediate dataset to inform wider regional healthcare strategies.

Through this feature selection process, we aim to construct a model that embodies the delicate balance between the inclusivity of relevant data and the exclusivity of noise, thereby maximizing predictive accuracy and reliability in real-world emergency triage situations. To incorporate categorical variables into machine learning models, they must be transformed into a numerical format. This involved:

- **Ordinal Encoding**: For categories with inherent order, ordinal encoding assigns each category a unique number, preserving the hierarchical structure important for predictions. This is theoretically represented as a mapping function $f : C \to \mathbb{R}$, where each category c_i in the set C is given a unique ordinal number.

- **Numeric Feature Handling**: Numeric features are kept as-is to maintain their original scale and distribution, avoiding potential biases from data normalization.
- **Target Variable Encoding**: The categorical outcome variable remains in its categorical form to keep classification outcomes clear and prevent introducing artificial order.

In essence, these methods ensure the variables retain their original meaning and relevance, crucial for creating interpretable and accurate models. An example visualization of the dataset after transformation is shown in Figure 5.

	Municipality of Residence	Age	Arrival Mode	Main Problem (1)	Main Problem (2)	Appropriateness Outcome
1	18	3	2	5	5	GREEN
2	18	3	2	5	5	GREEN
3	18	3	2	5	5	GREEN
4	7	27	2	9	9	GREEN
5	7	27	2	9	9	GREEN
6	7	27	2	9	9	GREEN
7	41	31	2	0	0	GREEN
8	41	31	2	0	0	GREEN
9	53	22	1	2	2	ORANGE
10	53	22	1	2	2	ORANGE
11	53	22	1	2	2	ORANGE

Figure 5: Emergency dataset after the transformation

4. Results and Discussion

4.1. Experimental Models

In the experimental phase, we tested various ML models for classification and pattern recognition:

In our study, we explored a variety of ML models for classification tasks. Random Forest constructs multiple decision trees to enhance prediction accuracy through ensemble techniques [11]. Naive Bayes utilizes Bayes' theorem with strong independence assumptions, proving efficient in handling high-dimensional datasets [12]. Neural Networks model complex relationships through interconnected neurons [13]. CN2 Rule Induction generates interpretable if-then rules [14]. AdaBoost improves classification strength by sequentially focusing on misclassified instances [15]. SVM operates effectively in high-dimensional spaces by identifying optimal separating hyperplanes [16]. Lastly, Stochastic Gradient Descent facilitates efficient optimization in large-scale learning problems [17]. We assessed our models' performance using Confusion Matrices and Lift Curves, combining them with cross-validation (k=5) to establish a comprehensive framework for evaluating predictive accuracy and model behavior on the dataset, as detailed in Table 2.

Table 2
Classification Test Results for Emergency Prediction

Model	AUC	CA	F1	Precision	Recall
kNN	0.972	0.869	0.866	0.864	0.869
Tree	0.972	0.960	0.959	0.959	0.960
SVM	0.885	0.741	0.658	0.710	0.741
SGD	0.498	0.704	0.606	0.538	0.704
Random Forest	**0.992**	0.957	0.956	0.955	0.957
Neural Network	0.863	0.754	0.671	0.634	0.754
Naive Bayes	0.818	0.694	0.693	0.703	0.694
CN2 Rule Induction	0.979	0.911	0.914	0.918	0.911
AdaBoost	0.986	**0.976**	0.977	0.979	0.976

Random Forest, CN2 Rule Induction, and AdaBoost demonstrated superior performance in emergency prediction tests, with Random Forest achieving an AUC of 0.992. These models showed high precision and recall, effectively identifying true positives while minimizing false positives and negatives, indicating their potential utility in clinical decision support systems. The Random Forest and AdaBoost models showcase exemplary performance with classification accuracies of 0.997 and 0.988 respectively, indicating their robustness and reliability in predicting emergency states. as highlighted in Figure 6.

Their confusion matrices are presented in Figure 7. The confusion matrices for AdaBoost and Random

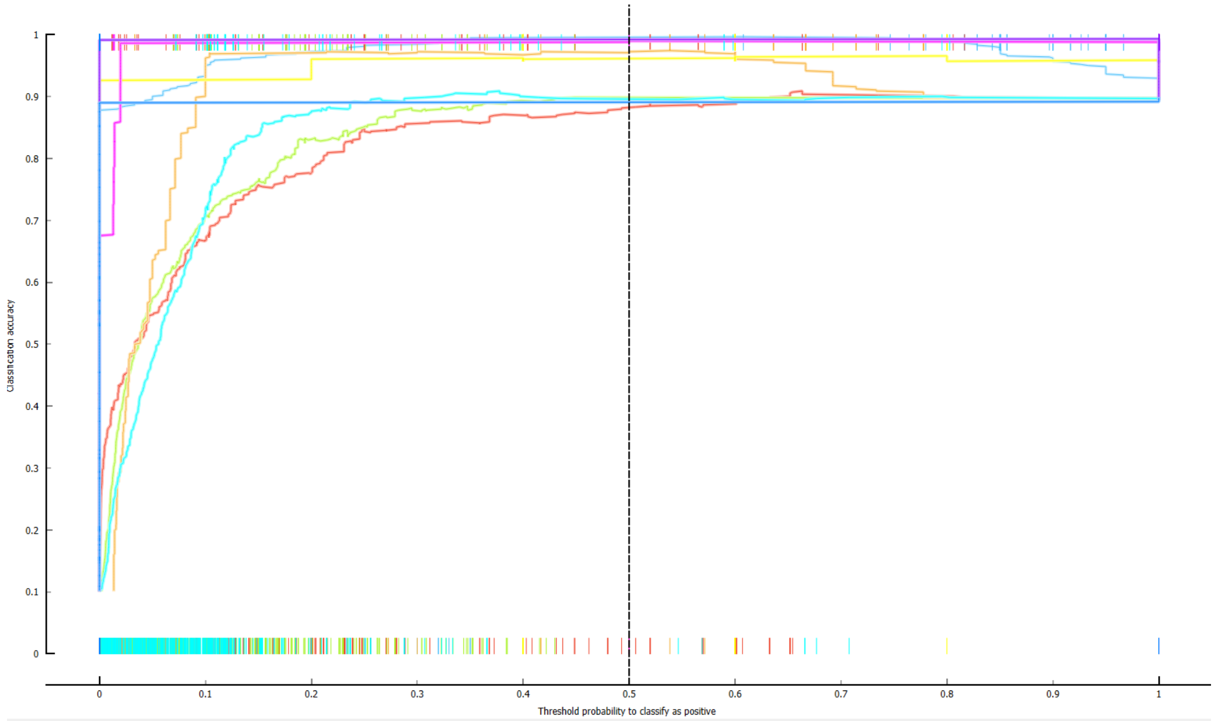

Figure 6: Calibration plot

Tree (Figure 7 indicate high accuracy in true positive rates, particularly in the ORANGE, BLUE, and RED classes for AdaBoost, and in the ORANGE and RED classes for Random Tree, evidencing their strength in correctly classifying the majority of cases.

The Tree algorithm's decision structure (Figure 8) reveals key paths to identifying the most complex RED cases: one significant pathway emerges when the Age is greater than 39 and the Main Problem is identified as critical, leading to a substantial proportion of RED outcomes. Another notable route is for patients over 73 years old, where the Main Problem significantly influences the likelihood of a RED classification, demonstrating the algorithm's capacity to discern intricate patterns that signal high emergency levels.

Using SHAP values (Fig. 9) and Random Forest feature importance analysis, Figure 10, reveals the impact of various features on emergency prediction, highlighting the significant roles of Age, Main Problem (1), and Main Problem (2). While showing Municipality of Residence and Arrival Mode as secondary and variable factors, respectively. This comprehensive analysis aids in refining ER triage and intervention strategies.

Age and Main Problem are identified as primary determinants of emergency case urgency, with Age exhibiting varied impacts and Main Problem a consistent, significant influence. Municipality of Residence and Arrival Mode offer additional context but exert a lesser and less variable influence on model predictions. In Fig. 11, the graphical representation of the Random Forest model shows terminal nodes representing the most critical (red) and non-critical (blue) cases at the extremes of decision tree branches, demonstrating the model's effectiveness in differentiating emergency levels, vital for ER decision-making and resource allocation.

Adaboost

		Predicted					
		ORANGE	BLUE	WHITE	RED	GREEN	Σ
Actual	ORANGE	96.7 %	0.0 %	0.0 %	0.0 %	3.3 %	60
	BLUE	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	40
	WHITE	0.0 %	0.0 %	100.0 %	0.0 %	0.0 %	21
	RED	0.0 %	0.0 %	0.0 %	100.0 %	0.0 %	38
	GREEN	0.5 %	0.7 %	1.7 %	0.0 %	97.2 %	423
	Σ	60	43	28	38	413	582

Random Tree

		Predicted					
		ORANGE	BLUE	WHITE	RED	GREEN	Σ
Actual	ORANGE	96.7 %	0.0 %	0.0 %	0.0 %	3.3 %	60
	BLUE	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	40
	WHITE	0.0 %	0.0 %	57.1 %	0.0 %	42.9 %	21
	RED	0.0 %	0.0 %	0.0 %	100.0 %	0.0 %	38
	GREEN	1.2 %	0.7 %	0.7 %	0.0 %	97.4 %	423
	Σ	63	43	15	38	423	582

Figure 7: Confusion Matrix for the best models in the experiment (Adaboost and Random Forest)

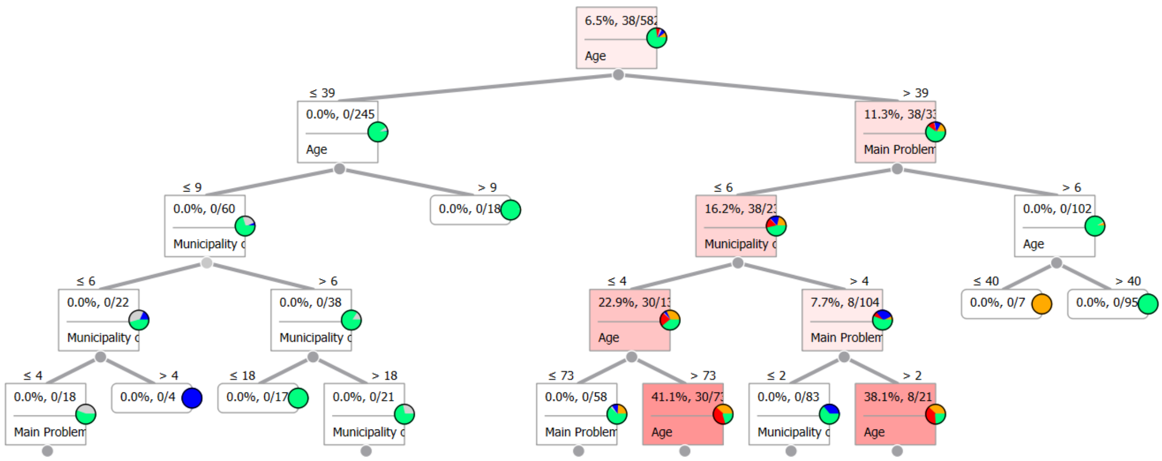

Figure 8: Solution generated by Tree algorithm to highlight the classification red.

5. Conclusion

This paper proposed the analysis and development of a computerized system capable of predicting emergencies in ERs based on data from a hospital in Northern Italy. Our methodology encompassed data preprocessing, feature engineering, and the application of machine learning models which delivered satisfactory and interpretable results. The high predictive performance of our models ensured relevant

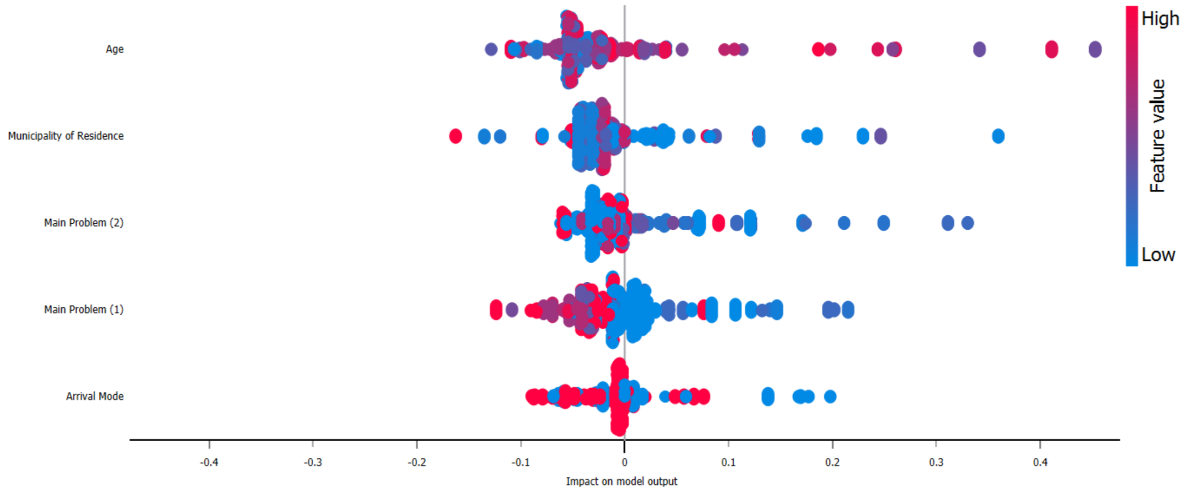

Figure 9: SHAP solution for Random Forest algorithm.

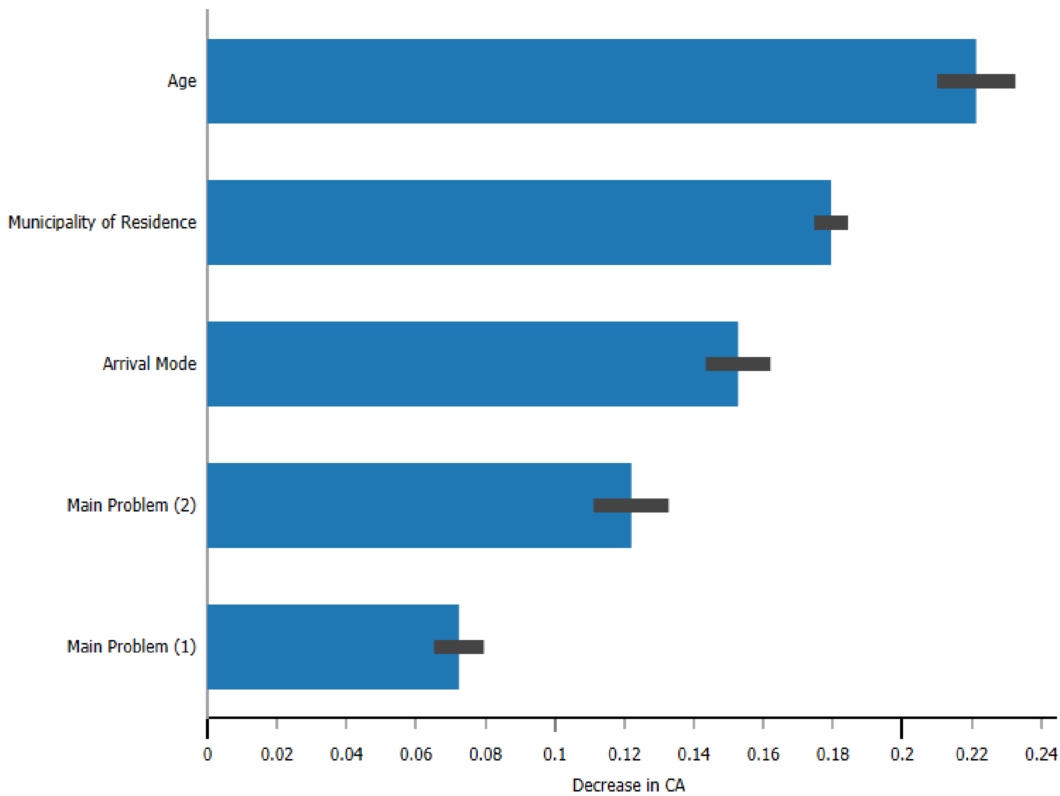

Figure 10: Feature Relevance for Random Forest algorithm.

insights into the nature of emergencies, offering valuable enhancements to the MTS protocol. Future work may extend predictions to other target variables, such as patient outcomes post-triage, to enable hospitals to better plan space and workforce allocation. The insights derived from our study affirm the transformative potential of AI in optimizing ER operations and patient care.

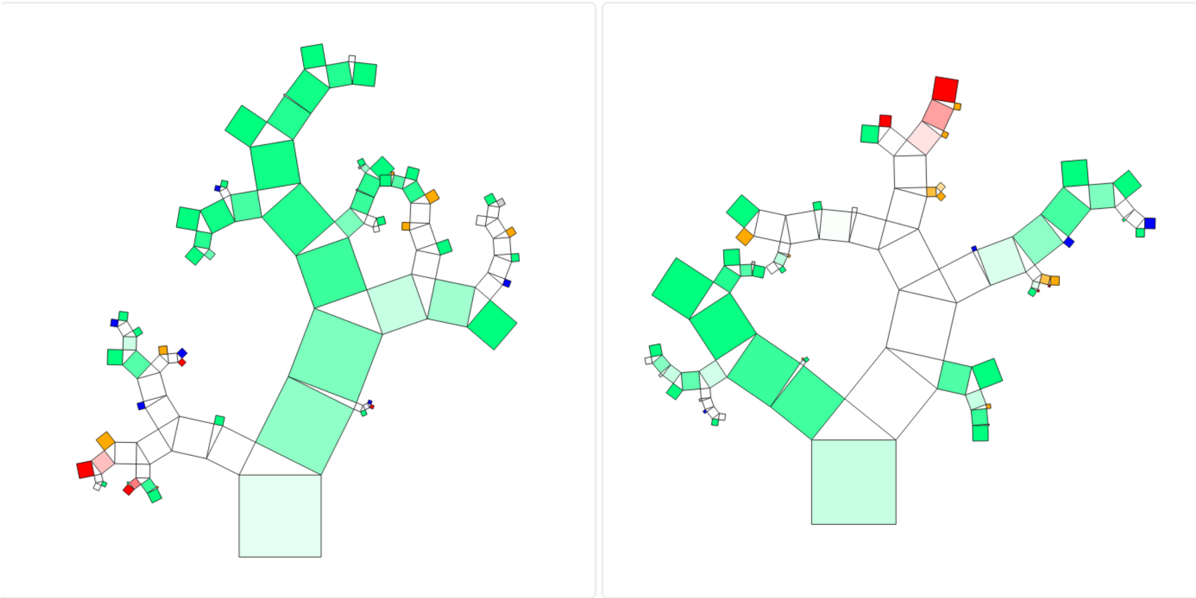

Figure 11: Random Forest- Graphical representation of the solution.

References

[1] J.-T. Lee, C.-C. Hsieh, C.-H. Lin, Y.-J. Lin, C.-Y. Kao, Prediction of hospitalization using artificial intelligence for urgent patients in the emergency department, Scientific Reports 11 (2021).

[2] H. Mutegeki, A. Nahabwe, J. Nakatumba-Nabende, G. Marvin, Interpretable machine learning-based triage for decision support in emergency care, in: 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), IEEE, 2023, pp. 983–990. doi:10.1109/ICOEI56765.2023.10125918.

[3] H. Martins, L. M. De Castro Dominguez Cuna, P. Freitas, Is manchester (mts) more than a triage system? a study of its association with mortality and admission to a large portuguese hospital, Emergency Medicine Journal 26 (2009) 183–186. doi:10.1136/emj.2008.060780.

[4] B. P. Roquette, H. Nagano, E. Marujo, A. C. Maiorano, Prediction of admission in pediatric emergency department with deep neural networks and triage textual data, Neural Networks 126 (2020) 170–177. doi:10.1016/j.neunet.2020.03.012.

[5] J. M. Zachariasse, J. Kuiper, M. de Hoog, H. Moll, M. van Veen, Safety of the manchester triage system to detect critically ill children at the emergency department, The Journal of Pediatrics 177 (2016) 232–237.e1.

[6] N. Seiger, M. van Veen, H. Almeida, E. Steyerberg, A. van Meurs, R. Carneiro, C. F. Alves, I. Maconochie, J. van der Lei, H. Moll, Improving the manchester triage system for pediatric emergency care: An international multicenter study, PLoS ONE 9 (2014).

[7] T. R. Azeredo, H. M. Guedes, R. A. Rebelo de Almeida, T. Chianca, J. Martins, Efficacy of the manchester triage system: A systematic review, International Emergency Nursing 23 (2015) 47–52.

[8] J. M. Zachariasse, N. Seiger, P. P. Rood, C. F. Alves, P. Freitas, F. Smit, G. Roukema, H. Moll, Validity of the manchester triage system in emergency care: A prospective observational study, PLoS ONE 12 (2017).

[9] A. P. Santos, P. Freitas, H. Martins, Manchester triage system version ii and resource utilisation in the emergency department, Emergency Medicine Journal 31 (2013) 148–152.

[10] Y. Li, T. Li, H. Liu, Recent advances in feature selection and its applications, Knowledge and Information Systems 53 (2017) 551–577.

[11] L. Breiman, Random forests, Machine learning 45 (2001) 5–32.

[12] G. H. John, P. Langley, Estimating continuous distributions in bayesian classifiers, Proceedings of the Eleventh conference on Uncertainty in artificial intelligence (1995) 338–345.

[13] W. S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, The bulletin of mathematical biophysics 5 (1943) 115–133.
[14] P. Clark, T. Niblett, The cn2 induction algorithm, Machine learning 3 (1989) 261–283.
[15] Y. Freund, R. E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of computer and system sciences 55 (1997) 119–139.
[16] C. Cortes, V. Vapnik, Support-vector networks, Machine learning 20 (1995) 273–297.
[17] L. Bottou, Large-scale machine learning with stochastic gradient descent, Proceedings of COMPSTAT'2010 (2010) 177–186.