SAYIŞTAY DERGİSİ

JOURNAL OF TURKISH COURT OF ACCOUNTS

Cilt/Volume: 36
Sayı/Issue: 139
Aralık/December 2025

ISSN: 1300-1981
eISSN: 2651-351X
Araştırma Makalesi/Research Article

A DATA-DRIVEN APPROACH TO THE ROLE CLASSIFICATION OF PUBLIC HOSPITALS: EVIDENCE FROM TÜRKİYE

KAMU HASTANELERİNİN ROL SINIFLANDIRMASINDA VERİYE DAYALI BİR YAKLAŞIM: TÜRKİYE'DEN KANITLAR

Tevfik BULUT[1]
Aziz KÜÇÜK[2]

ABSTRACT

Hospital classification plays a crucial role in modern healthcare systems by promoting efficient resource utilization, enhancing access to care, and directing patients to the most appropriate service levels. This study aimed to predict the role classifications of public hospitals in Türkiye using the Random Forest (RF) classification algorithm. It represents the first study conducted specifically on the role-based classification of public hospitals in Türkiye. The dataset included 716 public hospitals, categorized into eight distinct role classes. Two RF models (Model 1 and Model 2) were developed and evaluated using performance metrics, including overall accuracy, Cohen's Kappa coefficient, area under the curve (AUC), F1 score, and balanced accuracy. RF Model 2 consistently outperformed Model 1, achieving higher accuracy (96.82% vs. 95.82%), Kappa (0.9612 vs. 0.9489), and AUC (0.9889 vs. 0.9863). Thus, Model 2 is recommended for classifying the roles of public hospitals in Türkiye. The proposed approach can be adapted to classify hospitals in different healthcare systems, offering support for data-driven strategic planning and more equitable resource allocation.

1- PhD, Atılım University, Faculty of Health Science, buluttevfik@gmail.com, ORCID: 0000-0002-3668-7436.

2- Assoc. Prof. Dr., Ministry of Health, Directorate General for Public Hospitals, ORCID: 0000-0002-1296-4726.

Submitted/Gönderim: 26.07.2025, **Revision/Revizyon:** 09.11.2025, **Accepted/Kabul:** 09.11.2025

Corresponding Author/Sorumlu Yazar: Aziz Küçük, aziz.kucuk@yahoo.com

To Cite/Atıf: Bulut, T. and Küçük, A. (2025). A Data-Driven Approach to the Role Classification of Public Hospitals: Evidence from Türkiye. TCA Journal/Sayıştay Dergisi, 36 (139), 803-829. DOI: https://doi.org/10.52836/sayistay.1751816

ÖZ

Hastane sınıflandırması, modern sağlık sistemlerinde kaynakların etkin kullanımı, sağlık hizmetlerine erişimin artırılması ve hastaların uygun bakım düzeyine yönlendirilmesi açısından kritik bir rol oynamaktadır. Çalışmanın amacı, Türkiye'deki kamu hastanelerinin rol sınıflarını Rastgele Orman (Random Forest, RF) sınıflandırma algoritması kullanarak öngörmektir. Bu çalışma, Türkiye'de kamu hastanelerinin rol sınıflandırması üzerine doğrudan yürütülen ilk çalışmadır. Çalışmada, Türkiye genelinde sekiz farklı rol sınıfına ayrılmış 716 kamu hastanesine ait veriler analiz edilmiştir. İki farklı RF modeli (Model 1 ve Model 2) geliştirilmiş ve bu modeller genel doğruluk, Cohen'in Kappa katsayısı, eğri altı alan (AUC), F1 skoru ve dengelenmiş doğruluk gibi çeşitli performans göstergeleriyle karşılaştırmalı olarak değerlendirilmiştir. Yapılan performans değerlendirmesinde, RF Model 2'nin Model 1'e kıyasla üstün sonuçlar verdiği görülmüştür. Model 2, Model 1'e kıyasla daha yüksek doğruluk (%96,82'ye karşı %95,82), daha iyi Kappa katsayısı (0,9612'ye karşı 0,9489) ve daha yüksek AUC değeri (0,9889'a karşı 0,9863) elde etmiştir. Bu nedenle, Türkiye'deki kamu hastanelerinin rol sınıflarının öngörülmesinde Model 2'nin kullanılması önerilmektedir. Önerilen yöntem, farklı ülkelerdeki hastanelerin sınıflandırılmasında da uyarlanabilir ve sağlık yöneticilerine stratejik planlama ve kaynak tahsisinde veri temelli destek sağlayabilir.

Keywords: Public Hospital, Random Forest, Machine Learning, Hospital Classification, Türkiye.

Anahtar Kelimeler: Kamu Hastanesi, Rastgele Orman, Makine Öğrenmesi, Hastane Sınıflandırması, Türkiye.

INTRODUCTION

Recent advancements in artificial intelligence and machine learning (ML) have demonstrated substantial potential to reduce physicians' workload while simultaneously improving diagnostic accuracy, clinical predictability, and overall quality of care (Habehh and Gohel, 2021: 291). ML, a subfield of artificial intelligence, empowers algorithms to autonomously learn from data and generate predictions, thereby fostering a multitude of applications within the healthcare domain (An et al., 2023: 4178; Polat, 2024: 400). These encompass predictive analytics, wherein diverse datasets, including electronic health records, insurance claims, and genomic data, are leveraged to forecast patient outcomes, such as hospital readmissions or the onset of chronic diseases, facilitating proactive interventions. Diagnostic and therapeutic support is augmented by ML, particularly through techniques like convolutional neural networks, applied to medical imaging (e.g., CT scans, X-rays, MRIs) to enhance diagnostic precision and inform treatment strategies. Personalized medicine utilizes ML to integrate patient-specific characteristics (genetic predispositions, medical history, lifestyle factors) for predicting individual treatment responses and tailoring therapeutic regimens, critically relying on comprehensive genomic and longitudinal patient data. Clinical decision support systems incorporate ML to provide clinicians with real-time, data-driven insights, promoting informed decision-making and potentially improving patient outcomes. Examples include sepsis early-warning systems and risk stratification tools. Finally, population health management leverages ML to analyze large-scale population datasets, elucidating health trends, risk factors, and healthcare disparities. This analysis guides public health initiatives and optimizes resource allocation, as exemplified by targeted interventions for at-risk populations and disease outbreak prediction.

ML techniques aim to classify objects or predict values by applying algorithms and statistical analyses to large datasets (Köse and Polat, 2021: 17). Depending on the objective, ML can be categorized into three main areas: regression, supervised classification, and unsupervised classification (clustering) (De Oliveria et al., 2018: 1744). The ML category used in the study is supervised classification. This classification is a data analysis technique that utilizes classifiers to predict categorical class labels, allowing users to gain a deeper understanding of the data as a whole (Han et al., 2012: 330). In classification, mathematical functions map input variables (X) to output variables (Y) as targets, labels, or categories (Sarker, 2021).

The classification of hospitals is vital for the efficiency and sustainability of modern health systems. First and foremost, this classification ensures the optimal use of resources, enabling health services to reach a wider audience. Instead of each hospital having the same level of specialization and equipment, separating them according to different specialties and equipment needs allows for a more efficient distribution of medical devices, personnel, and budget. This not only prevents the wastage of resources but also enables better planning according to regional needs. Secondly, directing patients to the right hospitals improves treatment processes and makes more efficient use of vital time in emergencies. Classification allows patients to easily access hospitals with the level of specialization they need, preventing unnecessary waiting and incorrect treatment. Third, classification enables the establishment of distinct quality standards for various types of hospitals (Akdağ, 2012: 209; Övgün and Küçük, 2013). These standards facilitate performance monitoring and improvement efforts to enhance patient safety and quality of care. Finally, hospital classification also plays a vital role in identifying teaching hospitals and developing research infrastructure. In summary, hospital classification plays a crucial role in enhancing both the efficient use of resources and patients' access to the right services, as well as the overall quality of the health system.

In Türkiye, there is no study on predicting the role classes of hospitals using ML algorithms and other methods. This study constitutes the first nationwide, data-driven investigation directly focusing on the role-based classification of public hospitals in Türkiye. In the international literature, only a limited number of studies—primarily regional in scope, such as one conducted in the United States—have explored hospital classification using ML techniques. This study is driven by the need to address this critical gap in the literature. For the first time, it systematically applies a supervised ML approach, specifically the Random Forest (RF) classification algorithm, to predict the predefined role classes of public hospitals across Türkiye. The RF algorithm was selected due to its proven effectiveness in healthcare research, offering high predictive accuracy and the ability to model complex, nonlinear relationships between variables (Jackins et al., 2021; Simsekler et al., 2021; Song et al., 2021; Hong et al., 2022). The study also proposes an optimal model for role classification based on comprehensive performance metrics.

The study aims to predict public hospitals in Türkiye using RF models, and to identify the optimal RF model by comparing the obtained results with existing classification results based on performance metrics. Furthermore, the study aims to uncover the significance levels of the predictor variables utilized in role classification by the established RF models. Therefore, the goal is to identify the variables that need to be eliminated from the predictor variables of hospital role classes.

1. LITERATURE REVIEW

Hospital classification constitutes a fundamental component of health system organization, directly shaping the allocation of financial and human resources, access to healthcare services, referral pathways, and strategic planning processes. Globally, countries have adopted structured and context-specific frameworks to categorize hospitals based on functional roles, infrastructural capacity, service scope, and academic affiliations.

In the United States, hospital classification systems are multifaceted, often based on geographic location (e.g., rural or community hospitals), clinical specialization (e.g., women's, orthopedic, cardiac, or surgical hospitals), or institutional size. For instance, community-access hospitals are typically small rural facilities with fewer than 25 beds. In contrast, tertiary care or academic medical centers offer a wide range of specialties and subspecialties, including highly advanced fields such as pediatric cardiology (Griffin, 2012: 13–14). In contrast, countries such as China, Brazil, and Pakistan employ a standardized three-tiered hospital classification system, comprising primary, secondary, and tertiary levels, grounded in hierarchical service delivery. This model incorporates criteria such as bed capacity, service range, technological infrastructure, research activity, and educational functions. Primary-level hospitals are responsible for providing basic preventive and outpatient services at the community level, whereas tertiary-level institutions serve as regional referral centers equipped with comprehensive diagnostic, therapeutic, and academic capabilities (Li et al., 2022; Vieira- Meyer et al., 2022: 254; Jamal et al., 2024: 730). In Europe, particularly in the United Kingdom, Germany, and the Netherlands, hospital roles are integrated into national health systems through centralized planning and are explicitly aligned with referral hierarchies. Hospitals are generally designated as district hospitals, regional referral centers, or academic medical centers, based on their geographic coverage area, population served, medical specializations offered, and involvement in medical education and research (Healy & McKee, 2002: 59–74).

1.1. Turkish Public Hospitals by Role Classification

In Türkiye, the classification of hospitals is formally outlined in the 2005 Regulation on the Management of Inpatient Healthcare Facilities. According to this regulation, hospitals are defined as facilities that provide diagnostic, therapeutic, and rehabilitative services to both inpatients and outpatients, including maternity care services such as childbirth. The regulation categorizes hospitals into five functional groups: (1) district/town hospitals, (2) day hospitals, (3) general hospitals, (4) specialist hospitals, and (5) training and research hospitals (Tatar et al., 2011: 128).

Within the framework of the Health Transformation Program launched in 2003, regional planning efforts were initiated to optimize the allocation of healthcare resources and prevent the formation of idle capacity (Korkmaz and Tercan, 2023: 443). These efforts were shaped by Türkiye's geographical diversity, population distribution, distances from healthcare centers, regional transportation infrastructure, and existing health service inventories (Akdağ, 2012: 209–210). In line with this strategic planning approach, it became evident that the functional diversity and service scope of public hospitals required a more nuanced categorization beyond the formal typology defined in the 2005 Regulation on the Management of Inpatient Healthcare Facilities. Consequently, the need emerged to introduce a role-based classification model for public hospitals in Türkiye, aiming to align institutional functions with regional service demands, enhance system-level efficiency, and support evidence-based health planning (Atasever, 2021: 42). Thus, service delivery roles of existing and investment planning inpatient health facilities of the MoH were determined. Criteria for the Redefinition and Grouping of Roles of Ministry of Health Inpatient Health Facilities, which were put into practice with the Ministerial Decree dated December 3, 2009, and numbered 46143, were determined (Ministry of Health, 2011). In 2023, it underwent yet another update, taking into account the unique characteristics and needs of each group, the capacity of hospitals, their role in regional health planning, the number of specialized physicians, the availability of medical equipment, and the quality of their services (Ministry of Health, 2023). As presented in Table 1, public hospitals in Türkiye are categorized into eight distinct role classes: A1, A1 Branch, A2, A2 Branch, B, C, D, and E. The hospitals were analyzed within the framework of these designated role classifications.

Table 1: Identifying and Grouping the Roles of Hospitals

Roles	Definition and general characteristics
Group A1 Hospitals	These hospitals provide the most comprehensive and advanced health services in the country. They have the capacity to provide tertiary health care and are authorized for training and research.
Group A1 Branch Hospitals	They are designated as training and research hospitals, meeting specific criteria. These hospitals are advanced specialty hospitals that conduct training and research activities in their respective fields of specialization, possess high-tech equipment, and offer a Level III emergency service and intensive care unit.
Group A2 Hospitals	These hospitals offer a more limited range of services than those classified as Group A1 hospitals and lack authorization for training and research functions.
Group A2 Branch Hospitals	These are private specialty hospitals that are not authorized to conduct training and research activities.
Group B Hospitals	Located in provincial or fortified district centers, these hospitals provide less comprehensive services than Group A hospitals. They provide secondary health care services. These hospitals are situated in provincial centers or fortified district centers, distinct from Group A1 and A2 hospitals.
Group C Hospitals	These are smaller-scale hospitals typically located in fortified districts or sub-districts, characterized by a limited number of specialist physicians. They primarily provide primary and secondary healthcare services. Group C hospitals are generally smaller in size and offer a more limited range of medical specialties.
Group D Hospitals	These hospitals are located in fortified districts or sub-districts and have a minimum bed capacity of 25. They are designated to provide secondary healthcare services.
Group E Hospitals	These hospitals are typically located in small towns or sub-districts and have a maximum bed capacity of 25. They provide primary healthcare services and represent the smallest units within the healthcare system. Despite their limited capacity and specialization, Group E hospitals play a vital role, particularly in regions with restricted access to healthcare services.

Source: Ministry of Health (2023).

This hierarchical, role-based classification system ensures the efficient and equitable distribution of healthcare resources across Türkiye. It offers a range of services, from basic primary care in smaller communities to highly specialized tertiary care in major urban centers. The inter-tier referral system is of paramount importance in ensuring that patients receive the appropriate level of medical care.

1.2. Machine Learning Algorithms in the Healthcare

In recent years, the enhanced accessibility of extensive datasets and improvements in processing capabilities have facilitated the rapid development and dissemination of ML algorithms in healthcare. These algorithms have been employed to address a wide range of medical challenges, including medical imaging analysis, genetic disease prediction, personalized treatment planning, and outcome forecasting. Their application has demonstrated significant potential in enhancing diagnostic accuracy, optimizing treatment processes, and improving overall system efficiency.

Despite this growing body of research, the application of ML algorithms for hospital classification remains notably limited. One of the few studies in this area is by Dharmapala (2021: 22–29), who utilized multiple ML classification algorithms to categorize 45 hospitals in the United States into three groups: village, county, and regional hospitals. The classification was based on seven predictor variables: availability of drugs and equipment, occupancy rate, bed capacity, number of staff, average length of stay, number of inpatients, and number of live discharges. The study employed six different algorithms: decision trees, support vector machines, discriminant analysis, ensemble subspace k-nearest neighbor (KNN), Naive Bayes, and ensemble methods. Among these, the ensemble subspace KNN algorithm demonstrated the highest classification performance, as indicated by model validation and accuracy metrics.

Beyond hospital classification, RF—one of the most widely used tree-based ensemble learning algorithms—has been effectively applied to a broad range of predictive problems in healthcare. For example, Krämer et al. (2019:85-88) developed an RF-based model to classify hospital admissions as emergency or elective care based on patients' primary diagnoses. The model, trained using expert-annotated inpatient admission data, was also designed to assign a numeric urgency score to each diagnosis listed in the International Classification of Diseases catalog. Similarly, Simsekler et al. (2021) investigated the determinants of patient satisfaction across two key phases of the patient journey—registration and consultation. The authors constructed two separate RF models, incorporating both patient- and provider-related variables, using survey data collected from 411 patients at a hospital in the United Arab Emirates. The models aimed to estimate the relative importance of these variables in explaining satisfaction outcomes and to identify potential areas for improving healthcare quality. RF was preferred due to its high predictive accuracy and ability to model complex interactions among variables.

In another study, an interpretable RF model was developed to predict the severity of acute pancreatitis. The study retrospectively analyzed clinical and laboratory data from 648 patients, with missing values imputed using the multiple imputation by chained equations method. The RF model, trained in conjunction with a logistic regression model, outperformed its counterpart in predicting severe acute pancreatitis, demonstrating superior classification accuracy and robustness (Hong, 2022). The study by Jaotombo et al. (2022) aimed to identify the ML models that best predict prolonged hospital length of stay using a French medico-administrative database. The researchers used a retrospective cohort study design, collecting data from all discharges in 2015 from a large university hospital. They transformed length of stay into a binary variable (long vs. short, using the 90th percentile of 14 days as the cutoff). In the study, five ML models were applied: classification and regression trees, LR, RF, gradient boosting (GB), and neural networks (NN). Model performance was evaluated using AUC.

In a similar vein, Xie et al. (2022) sought to standardize predictive modeling practices in emergency department services by developing benchmark datasets and clinical prediction tasks from publicly available electronic health records. Their study introduced a unified preprocessing pipeline and evaluated a wide range of predictive algorithms, including LR, RF, GB, multilayer perceptron, Med2Vec, long short-term memory, and AutoScore, across multiple clinical outcomes. This facilitated cross-study comparison and advanced model development in emergency care. Likewise, Bhadouria and Singh (2023:27121) constructed predictive models for hospital length of stay and in-hospital mortality using the National Hospital Care Research Database (NHCRD), emphasizing the use of minimal feature sets to enhance interpretability and generalizability. Models were trained using several ML approaches, such as LR, RF, NB, GB, DT, ANN, bagging, and SVM, and evaluated with a comprehensive set of performance metrics (AUC, precision, recall, F-score, accuracy, MSE, MAE, RMSE). In another recent study, Choi et al. (2023) developed ML models to predict in-hospital fall risk among patients with acute stroke and compared their performance to the widely used Morse Fall Scale (MFS). Six algorithms-NB, KNN, RLR, SVM, RF, and XGB-were trained on EHR data, and several ML models outperformed the MFS in predictive accuracy, demonstrating the potential of data-driven approaches for improved fall risk assessment.

Collectively, these studies illustrate the growing utility of ML for predicting key hospital outcomes and strengthening clinical decision support. Notably, despite methodological variations, ensemble-based methods—particularly RF and GB—consistently demonstrate superior performance when applied to complex, multidimensional healthcare data.

2. MATERIAL AND METHODS

The target population of the study consists of 716 public hospitals in Türkiye classified into eight groups according to hospital role class. The distribution of the number of hospitals classified into eight groups according to role class is as follows: A1 (n = 82), A1 Branch (n = 21), A2 (n = 71), A1 Branch (n = 32), B (n = 133), C (n = 177), D (n = 110), and E (n = 90).

2.1. Data Set and Variables

The dataset comprises observations from public hospitals spanning the period from 2015 to 2023. All data were obtained with the approval of the Ministry of Health, Directorate General for Public Hospitals. Observations with missing values were excluded, and only complete cases were included in the RF models. The dataset was subsequently used to train and validate the RF models, ensuring robust predictive performance. Table 2 presents the variables included in the analysis. The study employs ten independent variables (predictors) to estimate the hospital role class as the dependent variable. Most variables in the dataset correspond to those currently used by the Ministry of Health in the role-based classification of hospitals. Additionally, the study incorporates predictor variables identified in the limited number of existing studies on hospital role classification (Dharmapala, 2021: 22-29; Tseng et al., 2015: 732), thereby contributing to the development of a more comprehensive predictive framework.

Table 2: Dependent and Independent Variables

Code	Variable	Data Type	Type of Variable
class	Hospital role class	Categorical 1. A1 2. A1 Branch 3. A2 4. A2 Branch 5. B 6. C 7. D 8. E	Dependent
tcys	Total number of hospital beds	Numerical	Independent
pms	Number of outpatient clinic visits	Numerical	Independent
hokg	Average length of stay	Numerical	Independent
tas	Total number of surgery procedures	Numerical	Independent
vki	Case mix index	Numerical	Independent
ths	Total number of physicians	Numerical	Independent

Code	Variable	Data Type	Type of Variable
hs	Total number of nurses	Numerical	Independent
tspds	Total number of non-health personnel	Numerical	Independent
tkdshs	Total number of referred patients	Numerical	Independent
ytdo	Bed occupancy rate	Numerical	Independent

2.2. R Packages for Analysis

The study utilized the R programming language (R Core Team, 2024) and RStudio (Posit Team, 2024) for data mining operations, data analysis, and result printing in Microsoft Excel workbooks. The packages used in the R environment are as follows: “caret” (Kuhn, 2008), “cvms” (Olsen, 2024), “doParallel” (Corporation and Weston, 2025), “dplyr” (Wickham et al., 2023), “ggplot2” (Wickham, 2016), “ggpubr” (Kassambara, 2022), “ggthemes” (Arnold, 2024), “here” (Müller, 2020), “openxlsx” (Schauberger and Walker, 2024), “pROC” (Robin et al., 2011), “readxl” (Wickham and Bryan, 2023), “skimr” (Waring et al., 2022), “tibble” (Müller and Wickham, 2023), “tidymodels” (Kuhn and Wickham, 2020), “tidyr” (Wickham et al., 2025), and “viridis” (Garnier et al., 2024).

2.3. RF algorithm for classification

To make predictions on new data, supervised ML techniques learn prediction rules from training data sets. ML algorithms can process a large number of predictors with nonlinear and complex interactions to find combinations of variables that reliably predict outcomes.

Random Forest for classification is an ensemble learning technique that builds several decision trees during training and produces the class that represents the mode of the classes (classification). This is an explanation of its functionality (Breiman, 2001: 5):

1) Bootstrap Aggregating (Bagging): The core of an RF is bagging. The training data is randomly sampled with replacement to create multiple subsets of the data (the same data point can appear multiple times in a single subgroup). Each subset is roughly the same size as the original training set.

2) Random Subspace: For each bootstrapped subset, a decision tree is grown. However, instead of considering all features (variables) at each split point in the tree, a random subset of features is selected. This is the “random subspace” method. The size of this subset is a hyperparameter that needs to be tuned.

3) Tree Growth: Each decision tree is grown to its full depth (no pruning) on its corresponding bootstrapped subset using the random subset of features. This is in contrast to other tree-based methods, where pruning is often used to prevent overfitting. The use of multiple trees with different splits compensates for the lack of pruning.

4) Classification: Once all the trees are grown, a new, unseen data point is fed into each tree. Each tree outputs a classification prediction (the class label). A majority vote determines the final classification. The class predicted by the majority of trees is the RF's prediction for the input data point.

RF combines bagging and random subspace to create a collection of diverse, unpruned decision trees. The final prediction is a consensus prediction from this eclectic collection.

2.4. Evaluation Metrics

The literature widely employs accuracy rate, Cohen's Kappa coefficient, F-score, and AUC classification metrics to evaluate the performance of RF and other ML models. One study used an RF algorithm to predict ten different diseases. In this study, accuracy, precision, sensitivity, recall, F score, and AUC metrics were used as classification metrics (Alam et al., 2019). Another study used a multilayer perceptron, RF, and LR methods to predict diabetes. This study used accuracy, precision, and recall metrics as classification metrics (Butt et al., 2021:1-5). The study used accuracy, AUC, and F-score classification metrics to predict the presence of diseases using artificial neural networks and LR methods (Bailly et al., 2022). Another study used six different ML algorithms to indicate diabetes. This study employed accuracy, Cohen's Kappa, precision, recall, F-score, and AUC as classification metrics (Adjei et al., 2022).

In this study, the balanced accuracy rate was used as a performance metric to evaluate hospitals according to their role classes. An overall accuracy rate is likely to yield erroneous results, especially in imbalanced datasets (Brodersen et al., 2010: 3121). In addition, the balanced accuracy metric minimizes the overall classification error (Thölke et al., 2023).

Table 3 presents the performance metrics along with their respective equations for assessing the models' performance. In performance metrics, TP denotes true positives, FP indicates false positives, FN represents false negatives, and TN signifies true negatives. Here, P0, which is used in the calculation of Cohen's Kappa coefficient, indicates the total probability of agreement or accuracy, and Pc suggests the likelihood of agreement due to chance (Ben-David, 2007: 874).

Table 3: Performance Metrics and Equations

Performance Metrics	Equations
Accuracy	(TP + TN)/(TP + TN + FP + FN)
AUC	$(TP_{rate} + TN_{rate})/2$
Balanced accuracy	(Sensitivity + Specificity)/2
Cohen's Kappa coefficient	$(P_0 - P_c)/(1 – P_c)$
F score	2 x (Precision x Recall) / (Precision + Recall)
Precision	TP/(TP + FP)
Recall	TP/(TP + FN)
Sensitivity	TP/(TP + FN)
Specificity	TN / (TN + TP)

AUC refers to the Receiver Operating Characteristic Area Under the Curve. AUC score provides a comprehensive measure of a classifier's performance across all classification thresholds. To obtain the score, it is necessary to calculate the area beneath the ROC curve. The AUC score indicates the classifier's ability to differentiate between positive and negative classes. The range of values is from 0.5 to 1. Higher values indicate better discrimination (De Hond et al., 2022: 853).

Cohen's Kappa was created to account for the probability that raters would guess on certain factors owing to uncertainty. The Kappa coefficient, like other correlation measures, can vary between -1 and +1 (Cohen, 1960: 37-46). The values for the Kappa coefficient in health research, as well as the degree of compliance, are as follows (McHugh, 2012): "0-0.20 (None)", "0.21-0.39 (Minimal)", "0.40-0.59 (Weak)", "0.60-0.79 (Moderate)", "0.80-0.90 (Strong)" and "Above 0.90 (Almost Perfect)".

2.5. Training and Testing

The stratified sampling method randomly selected observations from the dataset for the hospital role group to reduce variance between groups in the RF models. In the first model, we assigned 75% of the dataset (n = 3873) to the training dataset and 25% (n = 1291) to the testing dataset. Model 2 allocated 80% of the dataset (n = 4130) to the training dataset and 20% (n = 1034) to the testing dataset. Table 4 shows the training and testing sample sizes according to the models and hospital role class.

Table 4: Sample Sizes in the Training and Testing Sets by RF Model

Role	Model 1				Model 2			
	Training (%75)		Testing (%25)		Training (%80)		Testing (%20)	
	n	%	n	%	n	%	n	%
A1	457	11.8	152	11.77	487	11.79	122	11.8
A1Branch	107	2.76	36	2.79	114	2.76	29	2.8
A2	430	11.1	143	11.08	458	11.09	115	11.12
A2Branch	99	2.56	33	2.56	106	2.57	26	2.51
B	824	21.28	274	21.22	878	21.26	220	21.28
C	1096	28.3	365	28.27	1169	28.31	292	28.24
D	580	14.98	194	15.03	619	14.99	155	14.99
E	280	7.23	94	7.28	299	7.24	75	7.25
Total	**3873**	**100**	**1291**	**100**	**4130**	**100**	**1034**	**100**

2.6. Data Preprocessing

The “center” and “scale” parameters at this stage normalize the variables in the data set. In this context, the “center” parameter normalizes the data by subtracting the mean of the predictor variable observations from the predictor variable values. Conversely, the “scale” parameter divides the data of the predictor variable by its standard deviation.

2.7. Tuning

We set the tune length to 10 in both RF classification models. The “tuneLength” parameter in the Caret package in R indicates the number of different values to try in the models. The parameter returns the optimal “mtry” value that gives the highest accuracy rate as a result of the trials. The package utilizes the tuning parameter mtry to randomly select the number of variables in each split of the decision tree. The system determined the tuning mtry range values between 2 and 10. The models take the default value of 500 for the ntree parameter, which indicates the number of trees to grow. We did not exceed this value due to the possibility of overfitting.

Table 5 displays the mtry values that yielded the highest accuracy rates in both models. The tuning results indicate that Model 1 used the mtry parameter value of 4, which yields the highest accuracy rate. At this parameter value, the accuracy rate reaches approximately 83.9%. On the other hand, in Model 2, the mtry parameter value with the highest accuracy rate is 5, and the accuracy rate increases to 83.6% with this parameter value. The second model sets the mtry parameter value at 5.

Table 5: The Accuracy Rates in RF Models vary depending on the Parameter Values during Tuning

Model 1					Model 2				
mtry	Accuracy	Kappa	AccuracySD	KappaSD	mtry	Accuracy	Kappa	AccuracySD	KappaSD
2	0.8370	0.8000	0.0140	0.0172	2	0.8343	0.7967	0.0182	0.0224
3	0.8361	0.7990	0.0147	0.0183	3	0.8356	0.7984	0.0178	0.0219
4	0.8387	0.8023	0.0154	0.0191	4	0.8357	0.7986	0.0158	0.0193
5	0.8378	0.8011	0.0156	0.0193	5	0.8358	0.7989	0.0166	0.0204
6	0.8357	0.7986	0.0169	0.0210	6	0.8346	0.7973	0.0175	0.0214
7	0.8354	0.7983	0.0177	0.0218	7	0.8345	0.7972	0.0166	0.0204
8	0.8343	0.7969	0.0168	0.0208	8	0.8331	0.7955	0.0165	0.0202
9	0.8320	0.7940	0.0161	0.0200	9	0.8326	0.7949	0.0161	0.0198
10	0.8311	0.7929	0.0183	0.0227	10	0.8302	0.7919	0.0162	0.0199

SD: Standard Deviation

2.8. Model Validation

Model validation assesses the generalizability of a statistical analysis's results to an independent dataset (Seraj et al., 2023: 89). Various techniques are available for model validation. Cross-validation is a well-known and widely used method for model validation. Cross-validation (CV) comprises various data sampling techniques employed by algorithm developers to mitigate the risk of overoptimism in overfitted models. CV is used to assess the generalization performance of an algorithm and can additionally facilitate hyperparameter tuning and algorithm selection (Bradshaw et al., 2023). Cross-validation often yields less bias relative to holdout testing and has the distinct benefit of using the entire dataset for both training and testing. We used the repeated cross-validation method for validation in both RF models. This is because it is less biased than other methods in the literature. In a comparison of several cross-validation approaches on health data sets, 10-fold cross-validation without replication was distinctly inferior to the alternatives, exhibiting significantly larger absolute and mean squared errors (Smith et al., 2014: 319). The established models used k=10 and set the number of repetitions for the repeated cross-validation method to 3.

3. RESULTS

To further evaluate model robustness and investigate potential overfitting, we assessed the performance of both RF Model 1 and Model 2 on the whole dataset. The classification metrics for each model, stratified by hospital role class, are summarized in Table 6.

The overall accuracy of Model 1 on the whole dataset is approximately 95.82% (CI: 95.23%-96.35%) at a 95% confidence interval (CI). The Kappa coefficient is 0.9489, indicating almost perfect agreement between the hospital role classes in the entire dataset and the predicted role classes. Conversely, Model 2 achieves an overall accuracy rate of 96.82% (95% CI: 96.31%-97.29%). The Kappa coefficient in Model 2 is 0.9612, indicating a nearly perfect fit. Model 2 outperforms Model 1 when evaluated based on the overall accuracy rate and Kappa coefficient. Similarly, when hospital role classes evaluate the models using the F-score and balanced accuracy rate metrics, Model 2 outperforms Model 1 across all role classes.

Table 6: Results Obtained on the whole Dataset by the RF Model and Hospital Role Class

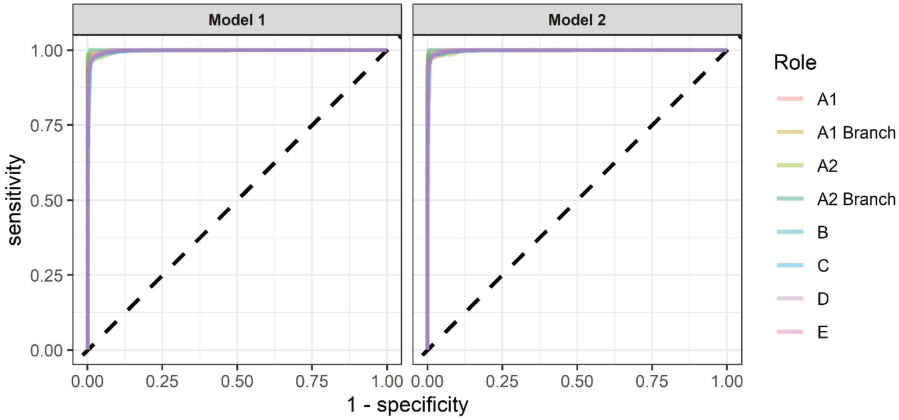
Model	Metric	A1	A1 Branch	A2	A2 Branch	B	C	D	E
Model 1	Sensitivity	0.969	0.951	0.951	0.939	0.959	0.973	0.951	0.917
	Specificity	0.998	1.000	0.994	0.999	0.988	0.982	0.989	0.997
	PPV*	0.988	0.986	0.954	0.961	0.956	0.956	0.940	0.961
	NPV**	0.996	0.999	0.994	0.998	0.989	0.989	0.991	0.994
	Precision	0.988	0.986	0.954	0.961	0.956	0.956	0.940	0.961
	Recall	0.969	0.951	0.951	0.939	0.959	0.973	0.951	0.917
	F Score	0.978	0.968	0.953	0.950	0.958	0.964	0.945	0.938
	Prevalence	0.118	0.028	0.111	0.026	0.213	0.283	0.150	0.072
	Balanced Accuracy	0.984	0.975	0.973	0.969	0.974	0.978	0.970	0.957

Model	Metric	A1	A1 Branch	A2	A2 Branch	B	C	D	E
	Sensitivity	0.979	0.951	0.960	0.955	0.972	0.977	0.961	0.944
	Specificity	0.999	1.000	0.996	0.999	0.990	0.986	0.992	0.998
	PPV*	0.990	1.000	0.970	0.969	0.963	0.965	0.958	0.970
	NPV**	0.997	0.999	0.995	0.999	0.992	0.991	0.993	0.996
Model 2	Precision	0.990	1.000	0.970	0.969	0.963	0.965	0.958	0.970
	Recall	0.979	0.951	0.960	0.955	0.972	0.977	0.961	0.944
	F Score	0.984	0.975	0.965	0.962	0.967	0.971	0.959	0.957
	Prevalence	0.118	0.028	0.111	0.026	0.213	0.283	0.150	0.072
	Balanced Accuracy	0.989	0.976	0.978	0.977	0.981	0.982	0.977	0.971

* **PPV:** Positive predictive value, ** **NPV:** Negative predictive value

Figure 1 displays the ROC curve produced by Models 1 and 2. Overall, Model 1 yielded an AUC value of 98.63%, while Model 2 yielded an AUC value of 98.89%, indicating near-perfect performance for both models. Therefore, a comparison of the two models reveals that Model 2 outperforms in terms of AUC value.

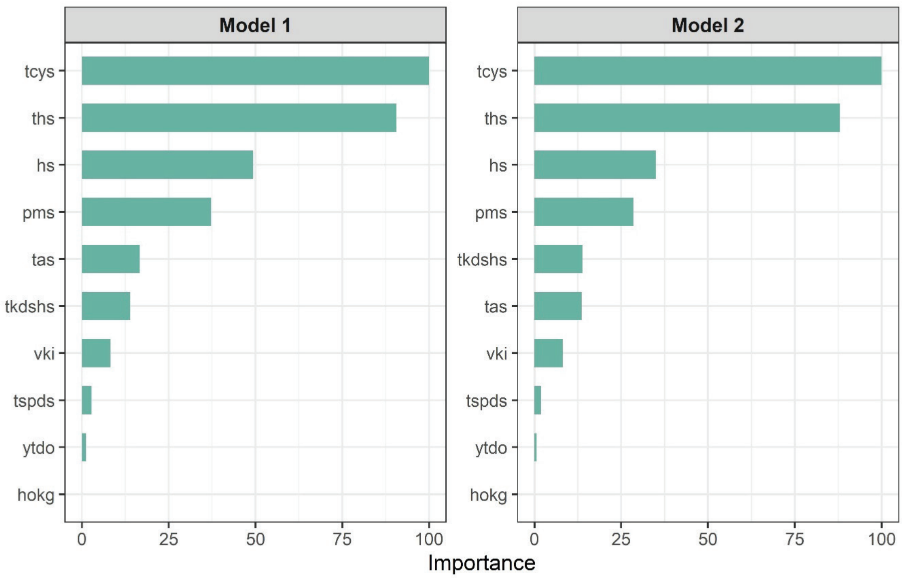
Figure 1: ROC Curve Obtained on the whole Dataset by the RF Model and Hospital Role Class



3.1. Importance Levels of Predictor Variables

After the RF models were trained and optimized, the importance scores of the predictor variables were calculated with the RF models and ranked from highest to lowest according to the model. Figure 2 illustrates the importance level of the predictor variables as determined by the RF models. The y-axis represents the predictor variables, and the x-axis represents the importance scores of the variables. In the models, "tcys" (total number of hospital beds), "ths" (total number of physicians), "hs" (total number of nurses), and "pms" (number of outpatient clinic visits) are the first four predictor variables with the highest level of importance in predicting the hospital role class, respectively. Similarly, the variables with the lowest significance levels in both models are "hokg" (average length of stay) and "ytdo" (bed occupancy rate).

Figure 2: Importance Levels of Predictor Variables by RF Model



DISCUSSION AND CONCLUSION

This study developed and evaluated RF classification models to predict the role classifications of public hospitals in Türkiye and to identify the relative importance of key predictor variables. The findings demonstrate that RF is a highly effective and transparent method for predicting hospital roles at the national scale. Among the models created, RF Model 2 demonstrated superior predictive performance based on multiple evaluation metrics and was thus recommended as the most accurate model. According to the model evaluation results, RF Model 2 consistently outperformed RF Model 1 in both overall classification accuracy and class-level metrics. Specifically, Model 2 achieved an overall accuracy of 96.82% and a Kappa coefficient of 0.9612, compared to 95.82% and 0.9489 for Model 1. The AUC values were also higher for Model 2 (0.9889) compared to Model 1 (0.9863). Furthermore, as shown in Table 6, Model 2 demonstrated better performance across all role classes in terms of F1 scores and balanced accuracy, supporting its robustness across diverse hospital types.

The strong predictive performance of RF aligns with the broader literature, which highlights its ability to handle complex nonlinear relationships and assess variable importance in health-related classification tasks (Krämer et al., 2019; Simsekler et al., 2021; Song et al., 2021; Hong et al., 2022). These characteristics make RF particularly suitable for national hospital classification systems, where organizational heterogeneity and multidimensional structural differences must be taken into account.

In Türkiye, there is no study on predicting the role class of hospitals using ML algorithms and other methods. Existing national research has generally focused on managerial or financial performance within predefined role groups (Keskin, 2018; Yılmaz & Şenel, 2019; Ekinci & Bakır, 2021; Koca & Demir Uslu, 2022; Babacan & Akça, 2024). In contrast, this study employs an algorithmic and data-driven classification method that offers a more objective and reproducible assessment framework. Internationally, the literature on hospital role classification using ML methods is also limited. The only comparable study was conducted in the United States by Dharmapala (2021), which classified 45 hospitals into three categories—village, county, and regional—using seven predictor variables and six ML algorithms. While the U.S. study achieved an accuracy of 93.3% using an ensemble subspace KNN model, its results were limited to a small regional sample. It used only accuracy and AUC as performance metrics. In contrast, the present study classified 716 public hospitals in Türkiye into eight distinct role categories (A1, A1-Branch, A2, A2-Branch, B, C, D, E), employed two RF models, and evaluated model performance using a broader set of metrics, including Kappa, F1 score, and balanced accuracy. Furthermore, this study provides class-specific model evaluation. It identifies key predictors of role classification, thereby offering more profound insights into the underlying structure of hospital roles at a national scale.

The identification of key predictive variables—such as bed capacity, staffing levels, and average length of stay—offers practical implications. These insights can support policymakers, strategic planners, and hospital administrators in designing more transparent, consistent, and evidence-based processes for role assignment and allocation. However, like most supervised learning techniques, the RF models could be further improved through expanded parameter tuning and the integration of additional predictors or alternative data sources. A remaining limitation is the exclusion of hospitals with incomplete data, underscoring the need for more standardized and comprehensive administrative datasets.

Although this study relies on RF as a supervised ML method, future research could explore alternative algorithms such as gradient boosting, SVMs, or deep learning models. Moreover, unsupervised ML approaches (e.g., clustering) could be used to uncover latent structures that may offer new or refined role groupings beyond the current administrative classification system. Such methods could contribute to more dynamic, adaptable hospital role frameworks that reflect the evolving complexity of healthcare delivery.

Overall, the results indicate that RF-based classification can support data-driven decision-making, improve transparency in role determination, and optimize resource allocation. The proposed modeling framework is adaptable and can be applied to other countries or health systems; however, its generalizability must be carefully considered. Because the model was trained on Türkiye-specific classification objectives, it cannot be applied to another context on a purely "plug-and-play" basis. Nonetheless, the methodological roadmap—objective variable selection, estimation of variable importance, and validation of administrative classification—offers a universal structure. Furthermore, health systems change over time, introducing the risk of model drift. Therefore, the model should be maintained using a dynamic MLOps approach, with continuous monitoring and periodic retraining based on up-to-date data.

In conclusion, RF-based hospital role prediction represents a promising tool for strengthening health system governance. In an era where healthcare systems are increasingly expected to "do more with less," data-driven decision support models, such as the one proposed in this study, can facilitate more informed and equitable healthcare planning.

REFERENCES

Adjei, H.P., Weyori, B., Boateng, O.N., Adekoya, A.F., Nti, I.K. (2023). Predicting Diabetes Using Cohen's Kappa Blending Ensemble Learning. International Journal of Electronic Healthcare, 13(1):1. https://doi.org/10.1504/ijeh.2023.10052670

Akdağ, R. (2012). Turkey Health Transformation Program Assessment Report (2003-2011), Ministry of Health of Turkey, Ankara.

Alam, Md. Z., Rahman, M.S., Rahman, M.S. (2019). A Random Forest based predictor for medical data classification using feature ranking. Informatics in Medicine Unlocked 15:100180. https://doi.org/10.1016/j.imu.2019.100180

An, Q., Rahman, S., Zhou, J., & Kang, J. J. (2023). A Comprehensive Review on Machine Learning in the Healthcare Industry: Classification, Restrictions, Opportunities, and Challenges. Sensors (Basel, Switzerland), 23(9), 4178. https://doi.org/10.3390/s23094178

Arnold, J. (2024). ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'. R package version 5.1.0.

Atasever, M. (2021). Sağlık Kurumları İşletmeciliği ve Hastane Yönetimi, Akademisyen Kitabevi: Ankara.

Babacan, A., & Akca, N. (2024). Kamu Hastanelerinin Finansal Performansının Oran Analizi ile Değerlendirilmesi: Türkiye Örneği. Muhasebe ve Finansman Dergisi, (101), 21-44. https://doi.org/10.25095/mufad.1313428

Bailly, A, Blanc, C, Francis, É., Guillotin, T, Jamal, F, Wakim, B, & Roy, P. (2022). Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models. Computer methods and programs in biomedicine, 213, 106504. https://doi.org/10.1016/j.cmpb.2021.106504

Ben-David, A. (2007). About the relationship between ROC curves and Cohen's kappa. Engineering Applications of Artificial Intelligence. 21(6), 874-882. https://doi.org/10.1016/j.engappai.2007.09.009

Bhadouria, A.S., Singh, R.K. (2024). Machine learning model for healthcare investments predicting the length of stay in a hospital & mortality rate. Multimed Tools Appl 83, 27121-27191. https://doi.org/10.1007/s11042-023-16474-8

Bradshaw, T.J., Huemann, Z., Hu, J., & Rahmim, A. (2023). A Guide to Cross-Validation for Artificial Intelligence in Medical Imaging. Radiology. Artificial intelligence, 5(4), e220232. https://doi.org/10.1148/ryai.220232

Breiman, L. (2001) Random Forests. Machine Learning, 45, 5-32. http://dx.doi.org/10.1023/A:1010933404324

Brodersen, K.H., Ong, C.S, Stephan, K.E and Buhmann, J.M. (2010). The Balanced Accuracy and Its Posterior Distribution. 2010 20th International Conference on Pattern Recognition, Istanbul, Türkiye, 3121-3124. https://doi.org/10.1109/ICPR.2010.764

Butt, U.M, Letchmunan, S, Ali, M, Hassan, F.H, Baqir, A, & Sherazi, H.H.R. (2021). Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications. Journal of healthcare engineering, 2021, 9930985. https://doi.org/10.1155/2021/9930985

Choi, J.H., Choi, E.S., & Park, D. (2023). In-hospital fall prediction using machine learning algorithms and the Morse fall scale in patients with acute stroke: a nested case-control study. BMC medical informatics and decision making, 23(1), 246. https://doi.org/10.1186/s12911-023-02330-0

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement. 20(1), 37-46. https://doi.org/10.1177/001316446002000104

Corporation, M, Weston, S. (2025). doParallel: Foreach Parallel Adaptor for the ‘parallel’ Package. R package version 1.0.17, https://github.com/revolutionanalytics/doparallel

De Hond, A.A.H., Steyerberg, E.W., & van Calster, B. (2022). Interpreting the area under the receiver operating characteristic curve. The Lancet. Digital health, 4(12), e853–e855. https://doi.org/10.1016/S2589-7500(22)00188-1

De Oliveira, H., Prodel, M., Augusto, V. (2018). Binary Classification on French Hospital Data: Benchmark of 7 Machine Learning Algorithms. 2018 IEEE International Conference on Systems, Man, and Cybernetics, 1743-1748. https://doi.org/10.1109/smc.2018.00301

Dharmapala, P. (2021). A Classification of Hospitals using Performance Features and Machine Learning Algorithms in the Event of Random Surge in Inpatients and Prediction of Live Discharges. Journal of Analytics 1(1):20-41. https://doi.org/10.29020/nybg.ja.v1i1.2

Ekinci, G, & Bakır, İ. (2021). Sağlık Kurumlarında Finansal Performans Analizi A1 Dal Hastanesi Örneği. Uluslararası Sağlık Yönetimi ve Stratejileri Araştırma Dergisi, 7(1), 1-18.

Garnier, S, Ross, N, Rudis R, Sciaini, M, Camargo, AP et al. (2024). viridis(Lite)-Colorblind-Friendly Color Maps for R. viridis package version 0.6.5. https://doi.org/10.5281/zenodo.4679423

Griffin, D.J. (Ed.) 2012, Hospitals: what they are and how they work, Fourth Edition, Jones & Bartlett Learning: Sudbury.

Habehh, H., & Gohel, S. (2021). Machine Learning in Healthcare. Current genomics, 22(4), 291–300. https://doi.org/10.2174/1389202922666210705124359

Han, J., Kamber, M., Pei, J. (2012). Classification. In: Data Mining, Third Edition, Elsevier: Waltham, MA. 327-391. https://doi.org/10.1016/b978-0-12-381479-1.00008-3

Healy, J. and McKee, M. (2002). The Role and Function of Hospitals. McKee, M. and Healy, J. (Eds). Hospitals in a Changing Europe. Open University Press, Maidenhead, pp. 59-82.

Hong, W., Lu, Y., Zhou, X., Jin, S., Pan, J., Lin, Q., Yang, S., Basharat, Z., Zippi, M., & Goyal, H. (2022). Usefulness of Random Forest Algorithm in Predicting Severe Acute Pancreatitis. Frontiers in cellular and infection microbiology, 12, 893294. https://doi.org/10.3389/fcimb.2022.893294

Jackins, V., Vimal, S., Kaliappan, M. et al. (2021). AI-based smart prediction of clinical disease using a random forest classifier and Naive Bayes. J Supercomput 77, 5198–5219. https://doi.org/10.1007/s11227-020-03481-x

Jamal, F., Ahmadini, A. A. H., Hassan, M. M., Sami, W., Ameeq, M., & Naeem, A. (2024). Exploring critical factors in referral systems at different health-care levels. World Medical & Health Policy, 16, 729–744. https://doi.org/10.1002/wmh3.632

Jaotombo, F., Pauly, V., Fond, G., Orleans, V., Auquier, P., Ghattas, B., & Boyer, L. (2022). Machine-learning prediction for hospital length of stay using a French medico-administrative database. Journal of market access & health policy, 11(1), 2149318. https://doi.org/10.1080/20016689.2022.2149318

Kassambara, A. (2022). Ggpubr: 'Ggplot2' Based Publication Ready Plots. https://cran.r-project.org/web/packages/ggpubr/index.html

Keskin, H.İ. (2018). Türkiye'de Sağlıkta Dönüşüm Programı ve Kamu Hastanelerinin Etkinliği. Akdeniz İİBF Dergisi, 18(38), 124-150. https://doi.org/10.25294/auiibfd.492741

Koca, M., and Demir Uslu, Y. (2022). Health Efficiency Scorecard Applications, Retrospective Evaluation of Hospitals' Efficiency with Stochastic Boundary Approach: The Case of AI Role Group Hospitals, Turkiye Klinikleri J Health Sci. 7(1): 212-20. https://doi.org/10.5336/healthsci.2021-81180

Korkmaz, E. ve Tercan, Ş. (2023). Sağlık Hizmetlerinin Sunumunda Yeni Bir Uygulama Alanı Olarak Medikal Muhasebe Üzerine Bir Araştırma. Sayıştay Dergisi, 34(130), 441-467. https://doi.org/10.52836/sayistay.1335177

Köse, H.Ö. and Polat, N. (2021), Dijital Dönüşüm ve Denetimin Geleceğine Etkisi, Sayıştay Dergisi, 32(123): 9-41 https://doi.org/10.52836/sayistay.1068328

Krämer, J., Schreyögg, J., & Busse, R. (2019). Classification of hospital admissions into emergency and elective care: a machine learning approach. Health care management science, 22(1), 85–105. https://doi.org/10.1007/s10729-017-9423-5

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. Journal of Statistical Software, 28(5), 1-26. https://doi.org/10.18637/jss.v028.i05

Kuhn, M., Wickham, H. (2020). Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. https://www.tidymodels.org

Li, L., Du, T., & Zeng, S. (2022). The Different Classification of Hospitals' Impact on Medical Outcomes of Patients in China. Frontiers in public health, 10, 855323. https://doi.org/10.3389/fpubh.2022.855323

McHugh, M.L. (2012). Interrater reliability: the kappa statistic. Biochemia Medica 276-282. https://doi.org/10.11613/bm.2012.031

Ministry of Health (2011). Health Services Requiring Specialized Planning in Türkiye 2011-2023. Publication No: 836, AG Design: Ankara.

Ministry of Health (2023). Ministry of Health Criteria for Determining the Roles of Inpatient Health Facilities. Date of the document: March 8, 2023. Number of the document: E-83913885. Ankara. https://www.saglik.gov.tr/TR-11024/saglik-bolge-planlamasi-hakkinda-genelge-ile-hastane-yatak-ve-rolleri-tescil-onayi-201050.html, 22.04.2025

Müller, K. (2020). Here: A Simpler Way to Find Your Files. R package version 1.0.1.

Müller, K, Wickham, H. (2023). tibble: Simple Data Frames. R package version 3.2.1, https://tibble.tidyverse.org/

Olsen, L., Zachariae, H. (2024). CVMS: Cross-Validation for Model Selection. R package version 1.6.2.

Övgün, B., and Küçük, A. (2013). Rescaling in Health Services: Regional Applications. Amme Idaresı Dergısı, 46(1), 57-80.

Polat, M. (2024). Yapay Zekânın Denetimde Kullanılması ve Etik Sorunlar. Sayıştay Dergisi, 35(134), 395-423. https://doi.org/10.52836/sayistay.1554497

Posit Team (2024). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. http://www.posit.co/

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12, 77. https://doi.org/10.1186/1471-2105-12-77

Sarker, I.H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. SN computer science, 2(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Schauberger, P, Walker, A. (2024). openxlsx: Read, Write and Edit xlsx Files. R package version 4.2.7.1, https://github.com/ycphs/openxlsx

Seraj, A, Mohammadi-Khanaposhtani M, Daneshfar R, Naseri M, Esmaeili M, Baghban A, et al. (2023). Cross-validation. In: Handbook of Hydroinformatics. Elsevier. p.89-105.

Simsekler, M. C. E., Alhashmi, N. H., Azar, E., King, N., Luqman, R. A. M. A., & Al Mulla, A. (2021). Exploring drivers of patient satisfaction using a random forest algorithm. BMC medical informatics and decision making, 21(1), 157. https://doi.org/10.1186/s12911-021-01519-5

Smith, G.C, Seaman, S.R, Wood, A.M, Royston, P, & White, I.R. (2014). Correcting for optimistic prediction in small data sets. American journal of epidemiology, 180(3), 318–324. https://doi.org/10.1093/aje/kwu140

Song, M., Jung, H., Lee, S., Kim, D., & Ahn, M. (2021). Diagnostic Classification and Biomarker Identification of Alzheimer's Disease with Random Forest Algorithm. Brain sciences, 11(4), 453. https://doi.org/10.3390/brainsci11040453

Tatar, M., Mollahaliloğlu, S., Şahin, B., Aydın, S., Maresso, A., HernándezQuevedo, C. (2011). Turkey: Health system review. Health Systems in Transition, 13(6):1–186.

Thölke, P., Mantilla-Ramos, Y. J., Abdelhedi, H., Maschke, C., Dehgan, A., et al. (2023). Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. NeuroImage, 277, 120253. https://doi.org/10.1016/j.neuroimage.2023.120253

Tseng, S.F., Lee, T.S., & Deng, C.Y. (2015). Cluster analysis of medical service resources at district hospitals in Taiwan, 2007–2011. Journal of the Chinese Medical Association, 78(12), 732-745. https://doi.org/10.1016/j.jcma.2015.05.013

Vieira-Meyer, A.P.G.F., Coutinho, M.B., Santos, H.P.G., Saintrain, M.V., & Candeiro, G.T.M. (2022). Brazilian Primary and Secondary Public Oral Health Attention: Are Dentists Ready to Face the COVID-19 Pandemic?. Disaster medicine and public health preparedness, 16(1), 254-261. https://doi.org/10.1017/dmp.2020.342

Waring, E, Quinn, M, McNamara, A, Arino de la Rubia E, Zhu H, Ellis S. (2022). skimr: Compact and Flexible Summaries of Data. R package version 2.1.5, https://github.com/ropensci/skimr

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, https://doi.org/10.1007/978-3-319-24277-4

Wickham, H, François, R, Henry, L, Müller, K. and Vaughan, D. (2023). dplyr: A Grammar of Data Manipulation. https://dplyr.tidyverse.org

Wickham, H. and Bryan, J. (2023) Readxl: Read Excel Files R Package Version 1.4.3. https://CRAN.R-project.org/package=readxl

Wickham, H, Vaughan D, Girlich M (2025). tidyr: Tidy Messy Data. R package version 1.3.1.9000, https://github.com/tidyverse/tidyr

Xie, F., Zhou, J., Lee, J.W., Tan, M., Li, S., Rajnthern, L.S., Chee, M.L., Chakraborty, B., Wong, A. I., Dagan, A., Ong, M.E.H., Gao, F., & Liu, N. (2022). Benchmarking emergency department prediction models with machine learning and public electronic health records. Scientific data, 9(1), 658. https://doi.org/10.1038/s41597-022-01782-9

Yılmaz, F. & Şenel, İ.K. (2019). Sağlık Kurumlarının Etkinliklerinin Veri Zarflama Analizi ile Değerlendirilmesi. Sosyal Güvence (15), 63-88. https://doi.org/10.21441/sosyalguvence.600856

KAMU HASTANELERİNİN ROL SINIFLANDIRMASINDA VERİYE DAYALI BİR YAKLAŞIM: TÜRKİYE'DEN KANITLAR

Tevfik BULUT

Aziz KÜÇÜK

GENİŞLETİLMİŞ ÖZET

Hastane sınıflandırması, modern sağlık sistemlerinde kaynakların dengeli dağılımı ve etkin kullanımı, sağlık hizmetlerine erişimin artırılması ve hastaların uygun bakım düzeyine yönlendirilmesi açısından kritik bir rol oynamaktadır. Bu kapsamda, 2003 yılında başlatılan Sağlıkta Dönüşüm Programı çerçevesinde sağlık kaynaklarının etkin tahsisini sağlamak ve atıl kapasite oluşumunu önlemek amacıyla bölgesel planlama çalışmaları yürütülmüştür. Bölgesel planlamalar; Türkiye'nin coğrafi çeşitliliği, nüfus dağılımı, sağlık merkezlerine olan uzaklık, ulaşım altyapısı ve mevcut sağlık hizmeti envanteri göz önünde bulundurularak şekillendirilmiştir. Bu çerçevede, 2009 yılında kamu hastanelerinin hizmet sunum rolleri yeniden tanımlanmış ve hastaneler çeşitli kriterler doğrultusunda sekiz farklı rol sınıfına (A1, A1-Dal, A2, A2-Dal, B, C, D, E) ayrılmıştır.

Bu çalışmanın amacı, Türkiye'deki kamu hastanelerinin rol sınıflarını Rastgele Orman (Random Forest, RF) algoritmaları kullanarak tahmin etmek ve elde edilen sonuçları mevcut sınıflandırma ile karşılaştırarak en yüksek performansı gösteren modeli belirlemektir. Türkiye'de hastane rol sınıflarının makine öğrenmesi yöntemleriyle öngörülmesine yönelik literatürde herhangi bir ulusal düzeyde çalışma bulunmamaktadır. Bu yönüyle çalışma, kamu hastanelerinin rol temelli sınıflandırmasına odaklanan, veriye dayalı ilk kapsamlı analiz olarak öne çıkmaktadır.

Çalışmada, Sağlık Bakanlığına bağlı 716 kamu hastanesine ait 2015–2023 yılları arasındaki veriler kullanılmıştır. Hastane rol sınıfı bağımlı değişken olarak tanımlanırken, bu sınıfı tahmin etmek amacıyla 10 bağımsız (öngörücü) değişken modele dâhil edilmiştir. Veri işleme ve analiz süreçleri ile sonuçların dışa aktarımı, R programlama dili ve R Studio yazılımı kullanılarak gerçekleştirilmiştir. RF modellerinin performansını artırmak amacıyla, "tuneLength" parametresi 10 olarak belirlenmiş ve "mtry" değeri optimize edilmiştir. Çalışmada iki farklı RF modeli (Model 1 ve Model 2) geliştirilmiş; bu modeller genel doğruluk, Cohen'in Kappa katsayısı, eğri altı alan (AUC), F1 skoru ve dengelenmiş doğruluk gibi çeşitli performans ölçütleri ile karşılaştırmalı olarak değerlendirilmiştir.

Çalışmanın bulguları, Model 2'nin genel performans açısından Model 1'e kıyasla üstün sonuçlar verdiğini göstermektedir. Bu üstünlük hem genel performansta hem de hastane rol sınıfları özelinde yapılan analizlerde geçerlidir. Model 2, %96,82 doğruluk, 0,9612 Kappa katsayısı ve 0,9889 AUC değeri ile öne çıkarken; tüm hastane rol sınıflarında daha yüksek F1 skoru ve dengelenmiş doğruluk sergilemiştir. Değişken önem analizine göre; toplam yatak sayısı, hekim sayısı, hemşire sayısı ve poliklinik başvuru sayısı en etkili öngörücü değişkenler olarak belirlenmiştir. Buna karşın, ortalama yatış süresi ve yatak doluluk oranının sınıflandırma üzerindeki etkisi oldukça düşük bulunmuştur. Bu doğrultuda, Türkiye'de kamu hastanelerinin rol sınıflarını tahmin etmek için en uygun modelin Model 2 olduğu sonucuna varılmıştır.

Elde edilen bulgular, RF algoritmasının hastane rol sınıflandırmasının yüksek doğruluk ve şeffaflıkla öngörülmesinde etkili bir yöntem olduğunu ortaya koymaktadır. Geliştirilen model, sadece sınıflandırma başarısıyla değil, aynı zamanda öngörücü değişkenlerin önem düzeylerini belirleyerek politika yapıcılara karar destek sunması açısından da değerlidir. Model, mevcut hastanelerin rollerinin yeniden değerlendirilmesi veya yeni kurulacak hastaneler için rol atamasının nesnel ve veri temelli biçimde yapılmasına olanak sağlamaktadır. Uluslararası literatürde, hastane sınıflandırmasını makine öğrenmesi teknikleriyle inceleyen çalışmalar oldukça sınırlıdır ve çoğunluğu, Amerika Birleşik Devletleri'nde gerçekleştirilen bir araştırma gibi, yerel düzeyde kalmaktadır. Bu bağlamda, önerilen yöntem yalnızca Türkiye'ye özgü bir katkı sunmakla kalmamakta; aynı zamanda farklı ülke ve sağlık sistemlerinde hastane sınıflandırmalarına uyarlanabilecek yapısıyla, kaynak tahsisi ve stratejik sağlık planlamasına yönelik küresel tartışmalara da katkı sağlamaktadır.