Kafkas Universitesi Veteriner Fakultesi Dergisi
Journal Home-Page: http://vetdergikafkas.org
E-ISSN: 1309-2251

Kafkas Univ Vet Fak Derg
30 (2): 233-241, 2024
DOI: 10.9775/kvfd.2023.30908

Research Article

Horse Surgery and Survival Prediction with Artificial Intelligence Models: Performance Comparison of Original, Imputed, Balanced, and Feature-Selected Datasets

Pınar CİHAN [1 (*)]

[1] Tekirdağ Namık Kemal University, Faculty of Corlu Engineering, Department of Computer Engineering, TR-59860 Tekirdağ - TÜRKİYE

(*) **Corresponding author:** Pınar CİHAN
Phone: +90 282 250 2456
E-mail: pkaya@nku.edu.tr

How to cite this article?
Cihan P: Horse surgery and survival prediction with artificial intelligence models: Performance comparison of original, imputed, balanced, and feature-selected datasets. *Kafkas Univ Vet Fak Derg,* 30 (2): 233-241, 2024.
DOI: 10.9775/kvfd.2023.30908

Article ID: KVFD-2023-30908
Received: 12.10.2023
Accepted: 03.01.2024
Published Online: 08.01.2024

Abstract

Artificial intelligence (AI) technology, while less advanced than in human medicine, holds significant potential in the field of veterinary medicine. This technology offers a range of essential benefits, such as disease diagnosis, treatment planning, disease control, and overall animal health improvement. Based on clinical data, this study uses 15 AI models to predict the necessity of surgery and the likelihood of survival in horses displaying symptoms of acute abdominal pain (colic). By comparing surgical and survival predictions across the original, imputed missing values, and balanced datasets, we determine the most effective dataset based on the average accuracy of the 15 AI models. Furthermore, we explore the potential for improved accuracy with a reduced feature set by calculating feature importance scores for surgery and survival predictions. Our results indicate that the balanced dataset achieved the highest average accuracy for predicting surgery and survival, with 80.76% and 77.96%, respectively. The Random Forest (RF) model outperformed others as the most accurate model for both surgery (accuracy = 85.83, Area Under the Curve [AUC] = 0.906) and survival prediction (accuracy = 80.75, AUC = 0.888). It was observed that reducing the number of features in the dataset by 56% led to an increase in surgery prediction accuracy to 86.38%. Similarly, when the number of features was reduced by 24% for survival prediction, the prediction performance increased to 83.75%. This study emphasizes the importance of the precise implementation of artificial intelligence techniques in veterinary medicine, which can significantly enhance model performance.

Keywords: Artificial intelligence, Data balancing, Feature selection, Horse colic, Prediction, SMOTE

Introduction

In recent years, research in artificial intelligence has gained significant attention and made notable advancements. The most popular tasks performed using AI models are prediction/classification [1,2] and regression problems [3,4]. However, in veterinary medicine, the use of this technology has not yet been as developed as in other areas. Nonetheless, studies conducted in veterinary medicine have demonstrated the potential for achieving successful results with this technology [5,6].

Horse health is one of the essential fields where artificial intelligence technology should be applied. This is because horses are valuable animals, often used in racing. Acute abdominal (colic) is one of the horses' most commonly encountered health issues. Colic presents a severe, health problem for horses and can result in death if not treated promptly. This issue is closely tied to a substantial economic impact and is a significant concern for horse owners [6]. Approximately 90% of colic cases in horses resolve spontaneously or with medical treatment, but the remaining 10% of colic cases can be fatal if not treated surgically [7]. Therefore, veterinarians strive to protect horses' health by recognizing colic symptoms and providing rapid intervention. At this point, artificial intelligence models come into play, offering fast and accurate predictions, and they can be used as a supportive tool in veterinarians' decision-making processes.

Artificial intelligence systems are based on mathematical models and require training using sample data. This training data includes the information that helps the model learn the desired outputs. Artificial intelligence is used to enhance computers' learning, decision-making, and problem-solving capabilities, and new models and applications are continuously being developed.

Artificial intelligence has been used in the field of veterinary medicine for body weight prediction [8], diagnostic radiology [9], blood sample value estimation [10], diagnosis of infectious and inflammatory disorders [11], classification of radiographs [12], animal diagnosis [13], milk yield prediction [14], and determination of bone fracture locations [15]. In a search conducted in the WOS database using the keywords "Horse colic + artificial intelligence," "Horse colic + machine learning," and "Horse colic + data mining," a relevant study was found. In this study, Fraiwan used artificial intelligence methods to predict with an accuracy rate of 76% and 85% the need for surgery and survival of horses with colic [6].

This study aims to make accurate surgery and survival predictions for -horse colic. For this purpose, the dataset used in the study contains missing values and is imbalanced. To address the difficulties posed by the dataset and achieve accurate predictions, the artificial intelligence methodology was meticulously implemented, and the prediction outcomes were evaluated at each stage.

The key contributions of this study are summarized as follows:

· This study demonstrates that accurately completing missing values in the dataset using the missForest method enhances prediction performance.

· This study addresses the issue of data imbalance by employing the SMOTE method and discovers that a balanced dataset enhances prediction performance.

· The study evaluates the performance of 15 AI models using accuracy and AUC statistical measures, and as a result, it indicates that the RF model outperforms other models.

· The study shows that a higher degree of surgical success and accuracy in predicting survival may be achieved by using fewer features.

Material and Methods

System Architecture

This study applied a series of steps to predict the need for surgical intervention and the probability of survival for horses with colic. The results obtained from the models may vary according to the dataset and the applied process steps [16]. Therefore, researchers can reduce or augment these steps according to the dataset and needs. The flowchart showing the classification process is shown in *Fig. 1*, and the subsequent sections provide a thorough explanation of each stage.

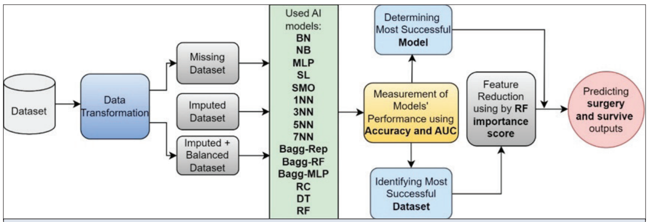

Fig 1. Flowchart for predicting of horse surgery and survival

Used Dataset

The dataset used in this study is publicly accessible online [17]. The dataset, containing medical data for 299 horses, consists of 27 features, with 25 used as inputs and two as outputs. The 'surgery' variable used as an output indicates whether the horses underwent surgery. The other output variable contains information about whether the horse survived. *Table 1* provides the characteristics of the features in the dataset, their types, feature information, missing value statistics, and information on using them as input/output in artificial intelligence models.

Data Transformation

Data transformation is the process of modifying or reshaping data to make it more suitable for dataset analysis or modeling [18]. This process improves data quality, ensures better model performance, and enhances data comprehensibility. Common data transformation techniques include feature scaling, normalization, transformation of categorical data, and handling outliers.

In this study, the dataset was transformed using categorical data transformation techniques. For example, the temperature of the extremities feature, with values normal, warm, cool, and cold, was transformed into 0, 1, 2, and 3, respectively. All categorical features underwent this transformation. Additionally, since the study aims to predict the survival status, the survival variable was assigned 0 for lived and 1 for dead and euthanized horse samples. The dataset used for predicting surgery and survival after data transformation was referred to as the *original dataset*.

Missing Value Imputation

In scientific research, researchers may encounter incomplete data collection, which may deviate from their anticipated parameters. Imputation is one of the popular data analysis approaches aimed at filling in missing data to make them usable [19]. The presence of missing values in the dataset presents a challenge for the planned analysis. The reason for this is that almost all classical and contemporary statistical techniques have been developed based on the assumption that the dataset is comprehensive [20]. In this study, Cihan [20] compared the accuracy performances of mean, kNN, SVD, bPca, and

Table 1. Characteristics and Information about the Horse Colic dataset

Feature	Type	Feature Information	#NA	Missing Rate	Direction
Age	Categoric	Adult, Young (<6 months)	0	0%	input
Temperature of extremities		Normal, warm, cool, cold	56	19%	
Peripheral pulse		Normal, increased, reduced, absent	69	23%	
Mucous membranes		normal pink, bright pink, pale pink, pale cyanotic, bright red, dark cyanotic	47	16%	
Capillary refill time		< 3 seconds, >= 3 seconds	32	11%	
Pain		No pain, depressed, mild pain, severe pain, extreme pain	55	18%	
Peristalsis		Hypermotile, normal, hypomotile, absent	44	15%	
Abdominal distension		None, slight, moderate, severe	56	19%	
Nasogastric tube		None, slight, significant	104	35%	
Nasogastric reflux		None, > 1 liter, < 1 liter	106	35%	
Rectal examination - feces		Normal, increased, decreased, absent	102	34%	
Abdomen		Normal, other, firm, small intestine, large intestine	118	39%	
Abdominocentesis appearance		Clear, cloudy, serosanguinous	165	55%	
Surgical lesion		No: Non-surgical lesion/Yes: Surgical lesion	0	0%	
Cp_data		No: Pathology data not present/Yes: data present	0	0%	
Rectal temperature	Numeric	Min: 35.4 - Max: 40.8	60	20%	
Nasogastric reflux ph		Min: 1 - Max: 7.5	246	82%	
Pulse		Min: 30 - Max: 184	24	8%	
Respiratory rate		Min: 8 - Max: 96	58	19%	
Packed cell volume		Min: 23 - Max: 75	29	10%	
Total protein		Min: 3.3 - Max: 89	33	11%	
Abdomcentesis total protein		Min: 0.1 - Max: 10.1	198	66%	
Lesion_1		Min: 0 - Max: 41110	0	0%	
Lesion_2		Min: 0 - Max: 7111	0	0%	
Lesion_3		Min: 0 - Max: 2209	0	0%	
Surgery	Categoric	No: horse had surgery/Yes: without surgery	0	0%	output
Outcome (Survive)	Categoric	Lived, Died, Euthanized	0	0%	output

missForest imputation methods. The analysis revealed that the missForest method successfully handled missing data in all different datasets. As a result, in this study, the missing values in the horseColic dataset were imputed using the missForest method.

The MissForest method utilizes Random Forest to predict missing values. It creates a separate Random Forest model for each missing variable. It uses these models to predict the missing values by considering the relationships between the missing variables, other complete variables, and the output variable [21]. This helps in accurately and reliably imputing missing data. The 'missForest' function in R programming was used to implement this method. This function completed the missing values in the dataset containing the 'surgery' output variable and then in the dataset containing the 'survive' output variable. In this study, the dataset version with missing values filled using the missForest method was referred to as the *imputed dataset*.

When applying artificial intelligence methods in WEKA, you can work with datasets that have missing values. WEKA ignores the missing data and does not use these instances in classification or artificial intelligence processes. Ignoring missing data results in a loss of information in the dataset. Therefore, it is essential to handle missing data properly.

Data Balancing

The situations where the output features (decision variables) in the dataset are not evenly distributed indicate that the dataset is imbalanced. Real-world datasets often imbalanced [22]. In imbalanced datasets, the class with a small number of samples is referred to as the minority class, while the class with a large number of samples is called the majority class [23]. Imbalanced datasets can mislead influence classification results, so it is desirable in artificial intelligence studies that the decision variable is evenly distributed in the datasets used. There are various methods that can be applied to eliminate imbalance in the preprocessing step of the data. The Synthetic Minority Over Sampling Technique (SMOTE) is one of the methods that can be used to address this issue, and it has been used in this study to address the data imbalance.

SMOTE is one of the most well-known and commonly used resampling methods. SMOTE creates new artificial minority class samples by interpolating among the existing minority class examples. This approach to generating synthetic samples was inspired by a technique used in handwritten character recognition. The method first finds the k nearest neighbors for each minority class example; then, it selects a random nearest neighbor. Subsequently, a new minority class example is created using the straight segment between a minority class example and its nearest neighbor. This process is repeated until both classes have an equal number of examples [24].

In this study, data balancing was performed using the SMOTE method after completing the missing values in the dataset. This dataset was referred to as a balanced dataset. For the survival output variable, the number of minority samples was increased by 47%, from 121 to 177, to balance the output sample size. Similarly, for the surgery output variable, the number of samples belonging to the minority class was increased by 48%, resulting in 180 'yes' and 180 'no' instances to balance the dataset.

Artificial Intelligence Models

In this study, colic horses' surgery and survival statuses were predicted using fifteen artificial intelligence classification methods. A list of the artificial intelligence classification algorithms used in this study, along with their brief descriptions, is presented in *Table 2*.

Table 2. *AI models used in this study*

Categories	Model	Abbr.	Description
Bayes	Bayes Network	BN	It is a probability theory-based graphical modeling approach. It uses a network structure that illustrates the relationships between variables and utilizes Bayes' theorem to make predictions [25].
	Naive Bayes	NB	Naive Bayes makes classifications based on the Bayes theorem and assumes independence between variables [26].
Functions	Multilayer Perceptron	MLP	It is one of the types of artificial neural networks. It makes complex decisions by mimicking the way the human brain works. It typically consists of input, hidden, and output layers.
	Simple Logistic	SL	It is a classification algorithm that fits the data with a suitable curve and has the ability to express class predictions of input features using probability distributions.
	Sequential Minimal Optimization	SMO	SMO is a classification algorithm based on Support Vector Machines (SVM). It maps the data into a high-dimensional space and provides a fit. It is developed to accelerate and optimize SVM training [27].
Lazy-learning algorithms	K-NN (K=1)	1NN	K-NN, to classify an instance, compares it with the classes of the K nearest neighbors. When a new data point arrives, K-NN calculates the K closest neighbors to this point, and by examining the classes of these neighbors, it determines the class of the new point. Since the choice of the K value can affect the classification results [28], different K values (1, 3, 5, 7) were tested in the study.
	K-NN (K=3)	3NN	
	K-NN (K=5)	5NN	
	K-NN (K=7)	7NN	
Meta-learning algorithms	Bagging REP Tree	Bagg-Rep	Bagging, each base classifier is trained independently and combines their results to create a stronger and more stable classifier. This can enhance prediction performance. Since each base classifier is trained on a different training subset, Bagging can reduce variance, thereby improving the model's prediction performance. In this study, prediction performances were tested using REP Tree, Random Forest, and Multilayer Perceptron as base classifiers.
	Bagging Random Forest	Bagg-RF	
	Bagging Multilayer Perceptron	Bagg-MLP	
	Random Committe	RC	It is a method that combines different classifiers to create a stronger classifier. First, it trains the dataset with random subsampling. Then, each classifier makes its own predictions. In the final step, the majority of these predictions are used to obtain the result [29]. In this study, Random Tree was used as the classifier.
Tree-based algorithms	Decision Tree	DT	A decision tree is represented as a tree structure, and each internal node is associated with a feature. As the dataset progresses through this tree structure, a decision is made at each node, and the data point is directed along a branch based on the relevant feature's value [13]. The C4.5 decision tree algorithm, known as J48 in WEKA, was used in this study.
	Random Forest	RF	It is an ensemble learning model that combines multiple decision trees. Each tree is trained on a random subset of the dataset, and this process is repeated randomly. Then, each tree classifies the data points, and the result of these classifications is subjected to majority voting [30].

This study used the R programming language [31] and the WEKA 3.8.5 (Waikato Knowledge Analysis Environment) tool [32]. The R programming language was employed for imputing missing data and determining feature importance scores. The WEKA tool was used for data balancing and classification tasks. The K-fold cross-validation (CV) technique was employed for both the training and testing phases of the classification methods. This technique divides the dataset into K equally sized subsets. While K-1 subsets are used for model training, the remaining subset is reserved for testing the model's performance. This process is repeated K times, with each subset used once for testing. The average of the K results is used to combine classification outcomes and assess overall performance. The CV technique is crucial for impartially evaluating model performance because it utilizes the entire dataset for model training and performance testing. In this study, a 10-fold cross-validation was implemented.

Measurement of Models Performance

The study used the Accuracy (ACC) and the Area Under the ROC Curve (AUC) metric to evaluate the prediction performance of artificial intelligence models for surgery and survival. Accuracy represents the ratio of correctly identified cases to the total number of cases (Equation 1). In this study, accuracy reflects the ability to correctly identify the need for surgical intervention and to accurately describe the horse's survival status.

$$Accuracy = \frac{Correct\ Predictions}{All\ Predictions} \quad (1)$$

The ROC curve is a graph that illustrates the variation in sensitivity and specificity at different threshold values, and the AUC measures the area under this curve. A high AUC value indicates that the model performs well and can classify accurately [33]. AUC was used to evaluate the ability to accurately predict the need for surgical intervention and the likelihood of a horse's survival.

Feature Selection

Feature selection is a process that involves choosing the most relevant and informative features from a dataset while excluding less important ones. By doing so, the dimensionality of the dataset is reduced, leading to faster AI model execution, lower memory consumption, and a streamlined analysis process. However, model performance may not always improve after the feature reduction process. While removing some features may enhance predictive performance, in some cases, it risks losing important information.

In fields such as veterinary medicine, data collection can be a challenging and costly process. The more accurate predictions can be made with less data, the more time and cost savings can be achieved. In this study, we calculated feature importance scores for the surgery and survival variables. We used the "importance" function in R programming for this purpose [34]. This method measures the importance of each feature and provides an importance ranking. Feature importance was calculated based on the Gini importance criterion, which indicates each feature's contribution to the model's predictive performance. After calculating the feature importance scores, we measured predictive performance by iteratively removing features, starting with the least important ones. This process identified the feature group that yielded the highest predictive performance.

Results

In this study, the prediction performances of AI models were analyzed after applying pre-processing and post-processing steps to achieve highly accurate predictions. In the pre-processing step, features in the dataset were transformed, missing values were handled, and data balancing procedures were carried out. In the post-processing step, a feature selection process was performed.

For the prediction of surgery and survival, a comparative performance analysis was conducted using 15 different artificial intelligence models on three different versions of the dataset. The original dataset containing missing data (referred to as original), the dataset in which missing values were imputed using the missForest method (referred to as imputed), and the dataset that was balanced after imputation (referred to as balanced) were compared in terms of the prediction performance of 15 AI models in *Table 3*.

According to the findings obtained, it was concluded that the 'balanced' dataset had the highest average accuracy for both surgery and survival predictions (80.76% and 77.96%, respectively). The dataset with the lowest average accuracy for both output variables was the original dataset (78.73% for surgery and 72.44% for survival).

According to the results presented in *Table 3*, the high accuracy value of the balanced dataset indicates that the processed and balanced version improved the accuracy of predictions. Therefore, the balanced dataset will be used for further procedures in the subsequent steps of the study. In evaluating the impact of datasets on prediction performance, using the average of 15 different AI models, rather than a single model, demonstrates that success is model-independent.

Furthermore, to determine the most successful model in making surgery and survival predictions, the prediction performance of 15 AI models was compared internally. The accuracy of surgery and survival predictions for the balanced dataset can be seen in *Table 3*, and the AUC values are presented in *Fig. 2*.

Table 3. Accuracy (%) results of models on different datasets for surgery and survival prediction						
Model	**Prediction of Surgery**			**Prediction of Survival**		
	Original Dataset	**Imputed Dataset**	**Balanced Dataset**	**Original Dataset**	**Imputed Dataset**	**Balanced Dataset**
BN	77.59	76.92	75.27	72.57	75.58	79.15
NB	71.57	73.24	73.33	73.57	75.58	78.02
MLP	77.25	76.92	78.61	68.89	70.90	73.23
SL	80.60	80.26	81.66	70.23	70.24	77.46
SMO	79.93	80.60	81.66	69.89	73.24	77.18
1NN	73.24	74.91	79.72	68.56	72.57	78.87
3NN	74.58	75.91	81.11	68.56	75.58	80.00
5NN	77.59	79.59	83.05	68.56	74.25	79.72
7NN	77.59	79.26	82.22	69.23	73.24	80.28
Bagg-Rep	80.94	80.6	81.94	73.57	74.58	76.34
Bagg-RF	82.60	82.94	84.16	78.26	79.26	80.56
Bagg-MLP	80.26	81.93	80.83	73.91	75.25	76.34
RC	83.27	78.59	82.50	76.92	73.91	80.00
DT	81.27	79.59	79.44	77.25	74.58	71.55
RF	82.6	84.28	**85.83**	76.58	73.25	**80.75**
Average	78.73	79.04	**80.76**	72.44	74.13	**77.96**

Feature selection was carried out to examine the impact of features in the dataset on prediction performance and, if possible, to make predictions with fewer features while achieving higher accuracy. The importance scores of features in the dataset were calculated using the *importance* function. In *Fig. 3* and *Fig. 4*, the importance scores of features for surgery and survival datasets and the prediction accuracy obtained through RF classification are presented.

When examining *Fig. 3*, it can be observed that the feature with the lowest importance for surgery prediction is lesion_3 (gini index = 0). Starting with the feature of lowest importance, features were iteratively removed from the dataset, and surgery prediction was made using the RF method. When all features were used, surgery was predicted with an accuracy of 85.83%, but when the 14 features with the lowest importance were removed from the dataset, the prediction accuracy increased to 86.38%.

As seen in *Fig. 4*, the accuracy rate of the survival prediction made using all features with the RF method is 80.75%. However, removing the least important six features from the dataset, increased the prediction accuracy to 83.85%. These results demonstrate that removing the

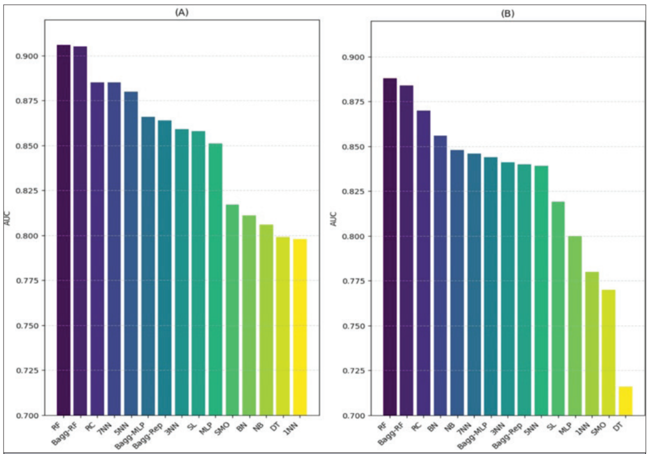

Fig 2. AUC results of 15 AI models for A- surgery and B- survival prediction

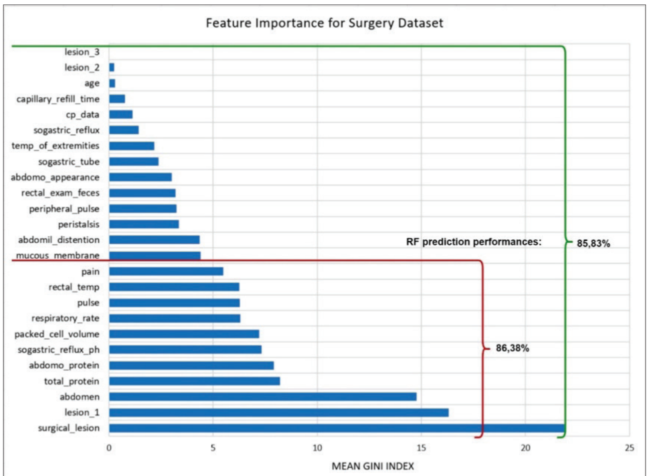

Fig 3. Impact of variables and RF prediction results on the Surgery Dataset

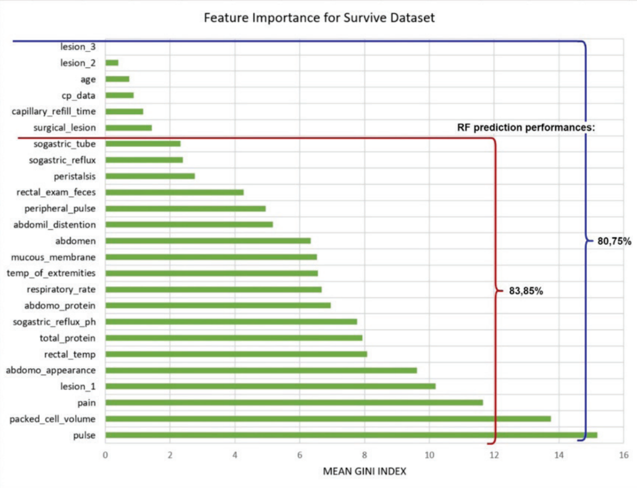

Fig 4. Impact of variables and RF prediction results on the Survive Dataset

least important features from the dataset enhances the performance of survival prediction.

These findings also indicate that feature selection improves the accuracy of surgery and survival predictions and that removing unimportant features can enhance model performance. The results also emphasize the importance of reducing model complexity to make more effective predictions in AI classification problems.

Discussion

In recent years, artificial intelligence methods have become widely used in various disciplines. However, in the field of veterinary medicine, the use of this technology has lagged behind and has yet to gain the expected popularity [6,35]. One of the primary reasons for the limited use of artificial intelligence in this field is that datasets are typically private and have limited accessibility. In human medicine, artificial intelligence has been extensively researched for many years, and there are many publicly available datasets. However, researchers often use data collected on animals in their own studies. The privacy of datasets hinders the development of interdisciplinary studies and restricts the transparency and accuracy of the obtained results [36]. We believe that sharing data publicly is a necessity for more interdisciplinary work in the field of veterinary medicine.

This study systematically applies artificial intelligence steps to the publicly accessible Horse Colic dataset on Kaggle, providing comparative results. Despite its publication many years ago, the Horse Colic dataset has yet to receive extensive study due to its complexity. The dataset includes data for 27 features from 299 horses, encompassing three types of data (continuous, discrete, and nominal). It also exhibits a relatively high rate of missing values *(Table 1)* and is imbalanced.

Missing values are a common issue in real-world dataset [37]. In some artificial intelligence applications, specific tools do not allow for model development on datasets containing missing data. In WEKA, artificial intelligence models can be used with datasets that include missing instances, but the missing samples are ignored. Both scenarios present challenges for researchers. Refrain from missing values further reducing the already limited number of examples in this field. Additionally, especially in medical datasets, accurately completing missing values is crucial to avoid making incorrect predictions. One of the primary objectives of this study is to fill in missing values in the dataset accurately. In many artificial intelligence applications, missing values in the dataset are either ignored or replaced with the mean value. However, several successful methods have been developed to complete missing values accurately. Mishandling missing values in a wrong or incomplete manner can lead to misleading or unreliable results. In this study, the missForest method, which has previously demonstrated its effectiveness, was used to impute missing values. The findings from the study indicate that the imputed dataset leads to more successful accuracy in surgery and survival predictions compared to predictions made with the original dataset.

The study's secondary aim is to deal with the challenge of imbalanced data. This is because dealing with dataset imbalances is important in improving prediction performance. Data imbalance issues arise in fields like veterinary medicine, where rare events like death are represented by a limited number of examples in the datasets. These imbalances negatively affect the model's learning process and prediction performance [38]. The SMOTE method was used in the study to address the dataset imbalance. When the prediction success of the balanced dataset is compared with the other datasets (original and imputed), the highest accuracy rates are obtained from the balanced dataset. These results demonstrate that carefully addressing data imbalance issues and using data balancing methods have a positive effect on the performance of model predictions.

The third main objective of the study is to identify the most effective artificial intelligence model for surgery and survival predictions. For this purpose, predictions were made using 15 different models. When comparing the performance of the models for surgery and survival predictions, it was concluded that the RF method is more successful than the other methods. The RF method has demonstrated successful results in solving various problems. Therefore, the RF method can be an effective option in various application areas in veterinary medicine.

The ultimate goal of the study is to make more successful predictions with fewer features in the dataset. This is because numerous dataset features can increase model complexity and raise the risk of overfitting. Therefore, feature selection makes the model more straightforward and effective. Additionally, since collecting medical data is a challenging process, making more accurate predictions with fewer features is important in data collection and resource utilization. In this study, feature importance scores were calculated using RF importance. It was concluded that when features with low importance were removed from the dataset, more accurate predictions could be made with fewer features *(Fig. 2)*. This is a crucial aspect to be considered in data analysis and artificial intelligence studies in veterinary medicine.

In conclusion, this study was conducted to explore the potential of artificial intelligence methods in veterinary medicine and address significant challenges in this area. Artificial intelligence holds great potential in veterinary medicine as well. However, its use in this field is still limited and faces essential barriers, such as the private and limited accessibility of datasets. The objective of this study was to enhance artificial intelligence research in veterinary medicine by showcasing methods to handle missing data, alleviate data imbalances, and simplify model complexity. In the future, creating and sharing more publicly accessible datasets could encourage the wider adoption of artificial intelligence methods in veterinary medicine. Additionally, interdisciplinary studies in this field could make valuable contributions to animal health and treatment.

Declarations

Availability of Data and Materials: The dataset used in the study is publicly available at: https://www.kaggle.com/datasets/uciml/horse-colic

Funding Support: This study was not financially supported.

Competing Interests: The author declared that there is no conflict of interest.

References

1. Cihan P, Ozger ZB: A new approach for determining SARS-CoV-2 epitopes using machine learning-based *in silico* methods. *Comput Biol Chem,* 98:107688, 2022. DOI: 10.1016/j.compbiolchem.2022.107688

2. Nasir N, Kansal A, Alshaltone O, Barneih F, Sameer M, Shanableh A, Al-Shamma'a A: Water quality classification using machine learning algorithms. *J Water Process Eng,* 48:102920, 2022. DOI: 10.1016/j.jwpe.2022.102920

3. Cihan P, Ozel H, Ozcan HK: Modeling of atmospheric particulate matters via artificial intelligence methods. *Environ Monit Assess,* 193, 1-15, 2021. DOI: 10.1007/s10661-021-09091-1

4. Tüfekci P: Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. *Int J Electr Power Energy Syst,* 60, 126-140, 2014. DOI: 10.1016/j.ijepes.2014.02.027

5. Vinicki K, Ferrari P, Belic M, Turk R: Using convolutional neural networks for determining reticulocyte percentage in cats. *arXiv:*180304873, 2018. DOI: 10.48550/arXiv.1803.04873

6. Fraiwan MA, Abutarbush SM: Using artificial intelligence to predict survivability likelihood and need for surgery in horses presented with acute abdomen (colic). *J Equine Vet Sci,* 90:102973, 2020. DOI: 10.1016/j.jevs.2020.102973

7. Hillyer M, Taylor F, French N: A cross-sectional study of colic in horses on Thoroughbred training premises in the British Isles in 1997. *Equine Vet J,* 33 (4): 380-385, 2001. DOI: 10.2746/042516401776249499

8. Topai M, Macit M: Prediction of body weight from body measurements in Morkaraman sheep. *J Appl Anim Res,* 25 (2): 97-100, 2004. DOI: 10.1080/09712119.2004.9706484

9. Müller TR, Solano M, Tsunemi MH: Accuracy of artificial intelligence software for the detection of confirmed pleural effusion in thoracic radiographs in dogs. *Vet Radiol Ultrasound,* 63 (5): 573-579, 2022. DOI: 10.1111/vru.13089

10. Cihan P, Gökçe E, Atakişi O, Kırmızıgül AH, Erdoğan HM: Prediction of immunoglobulin G in lambs with artificial intelligence methods. *Kafkas Univ Vet Fak Derg,* 27 (1): 21-27, 2021. DOI: 10.9775/kvfd.2020.24642

11. Ferrini S, Rollo C, Bellino C, Borriello G, Cagnotti G, Corona C, Di Muro G, Giacobini M, Iulini B, D'Angelo A: A novel machine learning-based web application for field identification of infectious and inflammatory disorders of the central nervous system in cattle. *J Vet Intern Med,* 37 (2): 766-773, 2023. DOI: 10.1111/jvim.16664

12. Costa da Silva RG, Mishra AP, Riggs CM, Doube M: Classification of racehorse limb radiographs using deep convolutional neural networks. *Vet Rec Open,* 10 (1):e55, 2023. DOI: 10.1002/vro2.55

13. Cihan P, Kalipsiz O, Gökçe E: Yenidoğan kuzularda bilgisayar destekli tanı. *Pamukkale Üniv Müh Bilim Derg,* 26 (2): 385-391, 2020. DOI: 10.5505/pajes.2019.51447

14. Topuz D: Lactation milk yield prediction with possibilistic logistic regression analysis. *Kafkas Univ Vet Fak Derg,* 27 (5): 547-557, 2021. DOI: 10.9775/kvfd.2020.25171

15. Baydan B, Barısçı N, Unver HM: Determining the location of tibial fracture of dog and cat using hybridized mask R-CNN architecture. *Kafkas Univ Vet Fak Derg,* 27 (3): 347-353, 2021. DOI: 10.9775/kvfd.2021.25486

16. Kumar Y, Koul A, Singla R, Ijaz MF: Artificial intelligence in disease diagnosis: A systematic literature review, synthesizing framework and future research agenda. *J Ambient Intell Humaniz Comput,* 14, 8459-8486, 2022. DOI: 10.1007/s12652-021-03612-z

17. Kaggle: Horse colic dataset: Can you predict the life or death of a horse?. https://www.kaggle.com/datasets/uciml/horse-colic; *Accessed:* 07.07.2023.

18. Kandel S, Heer J, Plaisant C, Kennedy J, Van Ham F, Riche NH, Weaver C, Lee B, Brodbeck D, Buono P: Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Inf Vis,* 10 (4): 271-288, 2011. DOI: 10.1177/14738716114159

19. Little RJ, Rubin DB: Statistical Analysis with Missing Data. John Wiley & Sons, 2019.

20. Cihan P: A comparison of five methods for missing value imputation in data sets. *Int Sci Voc Stud J,* 2 (2): 80-85, 2018.

21. Cihan P, Kalıpsız O, Gökce E: Veri madenciliği yöntemleriyle hayvan hastalıklarında teşhis, prognoz ve risk faktörlerinin belirlenmesi. *PhD Thesis*, Yıldız Teknik Univ., Fen Bil. Enst., 2018.

22. Cateni S, Colla V, Vannucci M: A method for resampling imbalanced datasets in binary classification tasks for real-world problems. *Neurocomputing,* 135, 32-41, 2014. DOI: 10.1016/j.neucom.2013.05.059

23. Saptono R, Winarno W, Drajati DP: Comparison of under-sampling and over-sampling techniques in diabetic mellitus (DM) patient data classification by using naive bayes classifier (NBC). *J Telecommun Electron Comput Eng,* 10, 145-148, 2018.

24. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP: SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res,* 16, 321-357, 2002. DOI: 10.1613/jair.953

25. Friedman N, Geiger D, Goldszmidt M: Bayesian network classifiers. *Mach Learn,* 29, 131-163, 1997. DOI: 10.1023/A:1007465528199

26. Sallam NM, Saleh AI, Arafat Ali H, Abdelsalam MM: An efficient strategy for blood diseases detection based on grey wolf optimization as feature selection and machine learning techniques. *Appl Sci,* 12 (21):10760, 2022. DOI: 10.3390/app122110760

27. Awad M, Khanna R, Awad M, Khanna R: Support vector machines for classification. In, Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. 39-66, Apress Open, 2015.

28. Sanlı T, Sıcakyüz Ç, Yüregir O: Comparison of the accuracy of classification algorithms on three data-sets in data mining: Example of 20 classes. *Int J Eng Sci Technol,* 12 (3): 81-89, 2020. DOI: 10.4314/ijest.v12i3.8

29. Niranjan A, Nutan D, Nitish A, Shenoy PD, Venugopal K: ERCR TV: Ensemble of random committee and random tree for efficient anomaly classification using voting. *2018 3rd International Conference for Convergence in Technology (I2CT). 06-08 April,* IEEE, 2018.

30. Breiman L: Random forests. *Mach Learn,* 45, 5-32, 2001. DOI: 10.1023/A:1010933404324

31. Venables W SD, R Core Team An Introduction to R-Version 3.0.2. 2013. http://cran.r-project.org/doc/manuals/R-intro.pdf; *Accessed:* 03.09.2023.

32. WEKA: Machine Learning Software in Java. https://www.cs.waikato.ac.nz/ml/weka/; *Accessed:* 05.09.2023.

33. Hajian-Tilaki K: Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J Intern Med,* 4 (2): 627-635, 2013.

34. LIAW A: Importance: Extract variable importance measure. 2020. https://www.rdocumentation.org/packages/randomForest/versions/4.6-14/topics/importance; *Accessed:* 03.09.2023.

35. Cihan P, Gokce E, Kalipsiz O: A review of machine learning applications in veterinary field. *Kafkas Univ Vet Fak Derg,* 23 (4): 673-680, 2017. DOI: 10.9775/kvfd.2016.17281

36. Appleby RB, Basran PS: Artificial intelligence in veterinary medicine. *J Am Vet Med Assoc,* 260 (8): 819-824, 2022. DOI: 10.2460/javma.22.03.0093

37. Tran CT, Zhang M, Andreae P, Xue B, Bui LT: Improving performance of classification on incomplete data using feature selection and clustering. *Appl Soft Comput,* 73, 848-861, 2018. DOI: 10.1016/j.asoc.2018.09.026

38. Ali H, Salleh MNM, Saedudin R, Hussain K, Mushtaq MF: Imbalance class problems in data mining: A review. *Indones J Electr Eng Comput Sci,* 14 (3): 1560-1571, 2019. DOI: 10.11591/ijeecs.v14.i3.pp1552-1563