

Evaluating the Efficacy of Bioelectrical Impedance Analysis Using Machine Learning Models for the classification of Parasitized Goats

[Aftab Siddique](#)*, [Phaneendra Batchu](#), Arshad Shaik, Priyanka Gurrapu, Tharun Tej Erukulla, Cornileus Ellington, Andrea L. Rubio Villa, Davia Brown, [Ajit Mahapatra](#), [Sudhanshu Panda](#), [Eric R Morgan](#), [Jan Van Wyk](#), [David Shapiro-Ilan](#), Govind Kannan, [Thomas H Terrill](#)

Posted Date: 25 November 2024

doi: 10.20944/preprints202411.1755.v1

Keywords: Bioelectrical Impedance; Machine Learning; Gastrointestinal Parasites; *Haemonchus contortus*; Veterinary Diagnostics



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Evaluating the Efficacy of Bioelectrical Impedance Analysis Using Machine Learning Models for the classification of Parasitized Goats

Aftab Siddique ^{1,*}, Phaneendra Batchu ¹, Arshad Shaik ¹, Priyanka Gurrapu ¹,
Tharun Tej Erukulla ¹, Cornileus Ellington ¹, Andrea L. Rubio Villa ¹, Davia Brown ¹,
Ajit Mahapatra ¹, Sudhanshu Panda ², Eric Morgan ³, Jan Van Wyk ⁴, David Shapiro-Ilan ⁵,
Govind Kannan ⁶ and Thomas Terrill ¹

¹ Department of Agricultural Sciences, Fort Valley State University, 1005 State University Drive, Fort Valley, GA 31030, USA

² Institute for Environmental Spatial Analysis, University of North Georgia, 3820 Mundy Mill Road, Oakwood, GA 30566, USA

³ Institute for Global Food Security, Queen's University, University Road, Belfast BT7 1NN, UK

⁴ Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Private Bag x04, Onderstepoort 0110, South Africa

⁵ United States Department of Agriculture- Agriculture Research Services, Fruit and Tree Nut Research, Byron, GA 31008, USA

⁶ Department of Poultry Sciences, 201 Poultry Science Building, Auburn Univ, AL 36801, USA

* Correspondence: aftab.siddique@fvsu.edu

Abstract: Rapid identification and assessment of animal health are critical for livestock productivity, especially for small ruminants like goats, which are highly susceptible to blood-feeding gastrointestinal nematodes, such as *Haemonchus contortus*. This study aimed to establish proof of concept for using bioelectrical impedance analysis (BIA) as a non-invasive diagnostic tool to distinguish parasite-infected goats from healthy ones. A cohort of 94 intact male Spanish goats (58 healthy; 36 parasitized; naturally infected with *Haemonchus contortus*) was selected to evaluate the efficacy of BIA through the measurement of resistance (Rs) and electrical reactance (Xc). Data were collected from live goats using the CQR 3.0 device over multiple time points. The study employed several machine learning models, including Support Vector Machines (SVM), Backpropagation Neural Networks (BPNN), k-Nearest Neighbors (K-NN), XGBoost, and Keras deep learning models to classify goats based on their bioelectrical properties. Among the classification models, SVM demonstrated the highest accuracy (95%) and F1-score (96%), while K-NN showed the lowest accuracy (90%). For regression tasks, BPNN outperformed other models with a nearly perfect R² value of 99.9% and a minimal mean squared error (MSE) of 1.25e-04, followed by SVR with an R² of 96.9%. The BIA data revealed significant differences in Rs and Xc between healthy and parasitized goats, with parasitized goats exhibiting elevated resistance values, likely due to dehydration and tissue changes caused by parasitic infection. These findings highlight the potential of BIA combined with machine learning to develop a scalable, rapid, and non-invasive diagnostic tool for monitoring small ruminant health, particularly in detecting parasitic infections like *Haemonchus contortus*. This approach could improve herd management, reduce productivity losses, and enhance animal welfare.

Keywords: Bioelectrical Impedance; Machine Learning; Gastrointestinal Parasites; *Haemonchus contortus*; Veterinary Diagnostics

1. Introduction

In warm and wet regions, small ruminants such as goats and sheep are especially vulnerable to parasitic infections, with *Haemonchus contortus* posing one of the greatest threats [1]. Commonly referred to as the barber pole worm due to its distinctive appearance, *Haemonchus contortus* thrives in warm, humid environments, infecting the abomasum (the fourth stomach chamber) of ruminants [2,3] by attaching to stomach lining and feeds on the host's blood, leading to substantial blood loss. Over time, infected animals develop anemia, which severely impacts their overall health, productivity, and even survival [2]. For farmers and livestock producers, infections from *Haemonchus contortus* represent a significant economic burden, causing diminished growth rates, lowered reproductive success, reduced milk production, and, in severe infestations, death.

Anemia in infected animals is a hallmark sign of *Haemonchus contortus* infection and is traditionally diagnosed using Hematocrit Analysis, a method that measures the packed cell volume (PCV) of blood [4,5]. The PCV analysis provides an estimate of the level of anemia, reflecting the degree to which the animal is suffering from blood loss. Although reliable, Hematocrit Analysis is labor-intensive, requiring invasive blood sampling and the expertise of trained personnel. This limits its scalability, especially for large herds, and introduces additional stress to the animals [6,7]. The cost and time involved in performing Hematocrit Analysis make it less practical for routine monitoring in large or resource-limited operations.

Another traditional approach used to diagnose parasitic infections in small ruminants is the fecal egg count (FEC). This method involves examining the number of parasite eggs present in an animal's feces to estimate the parasite burden within the host [8,9]. Fecal egg count is particularly useful for monitoring gastrointestinal nematodes such as *Haemonchus contortus*, as the number of eggs passed in the feces correlates with the worm burden in the animal [10,11]. The FEC method is commonly used in farm management programs to assess the need for treatment and to monitor the efficacy of deworming programs [12,13]. However, like Hematocrit Analysis, FEC is labor-intensive, requiring specialized equipment and expertise to collect, process, and interpret fecal samples [14,15]. Additionally, FEC results can vary significantly depending on the animal's diet, hydration status, and the time of day when samples are collected. Moreover, the accuracy of fecal egg counts may decrease when the parasite burden is low, leading to potential underestimation of the infection [16].

To address the limitations of traditional diagnostic methods such as PCV and FEC, researchers have been exploring alternative technologies that offer rapid, non-invasive, and scalable solutions. One such promising tool is bioelectrical impedance devices for analysis (BIA), a technique that measures the electrical properties of biological tissues, specifically electrical resistance (R) and electrical reactance (Xc) [17,18]. BIA works by passing a small, painless electrical impulse through the body and measuring how the tissues oppose the flow of the current [19–20]. These measurements taken through BIA provide insights into the body's composition, including water content, fat mass, and cellular health [21,22], all of which can be affected by parasitic infections.

In the case of *Haemonchus contortus*, the parasite's blood-feeding behavior disrupts the host's fluid balance and reduces the total volume of blood and red blood cells, which can alter the electrical properties of the tissue. This makes BIA a promising tool for detecting parasitic infections, as changes in the body's fluid levels and tissue composition are directly related to the severity of the infection. Unlike Hematocrit Analysis or FEC, BIA can be performed quickly, without the need for invasive procedures or specialized personnel. This makes BIA particularly well-suited for large-scale herd management, where rapid, non-invasive diagnostics are crucial for maintaining animal health and productivity.

Given the advantages of BIA in terms of ease, speed, and non-invasiveness, this study explored its potential in combination with machine learning (ML) techniques to classify goats as either parasitized or healthy. Machine learning algorithms, known for their ability to recognize complex patterns in large datasets, are applied to the bioelectrical impedance data collected from live goats. By training machine learning models on this data, the study aims to develop an efficient, scalable, and accurate diagnostic tool for detecting parasitic infections in small ruminants. The implications of this research go beyond the detection of *Haemonchus contortus*. The successful application of BIA and

machine learning could pave the way for diagnosing a wide range of health conditions in livestock, from other parasitic infections to metabolic disorders and nutritional deficiencies. By providing a rapid, non-invasive, and cost-effective diagnostic tool, this study lays the groundwork for the integration of advanced technologies in precision livestock farming, where early detection and proactive management are key to maximizing productivity and animal welfare. The following sections 1.1, 1.2, 1.3, 1.4 provides the basic working principle for the different techniques used in the study.

1.1. Support Vector Machines (SVM)

Support Vector Machines (SVMs) are a category of supervised machine learning methods employed for classification, regression, and anomaly detection. The SVM technique is widely recognized in machine learning for its efficacy in handling both linear and nonlinear classification challenges. The fundamental concept of SVM is to identify a hyperplane in an N-dimensional space (where N denotes the number of features) that effectively segregates data points into various classes [23,24]. Winston [25] compares this approach to “Fitting the widest possible street,” elucidating the quadratic optimization challenge associated with hyperplane separation using the equation [26]:

$$wx + b = 0 \dots\dots\dots (1)$$

w represents the weight vector, x is the input vector, and b signifies the bias factor. In SVM, hard-margin and soft-margin classifiers identify the optimal separation distances between classes, facilitating the classification of distinct patterns or features from an image [27].

In binary classification, numerous hyperplanes can delineate the data points of two classes, whereas in multi-class classification, more than two classes are taken into account. The goal is to identify the hyperplane that maximizes the margin, defined as the maximum distance between the support vectors data points nearest to the hyperplane. The support vectors are essential, as their removal would change the hyperplane's position. Consequently, SVMs are regarded as essential elements of the dataset. The efficacy of SVM resides in its capacity to manage nonlinear data using the kernel trick, which converts non-linearly separable data into a linearly separable format [28–30].

1.2. K-Nearest Neighbor (K-NN)

K-Nearest Neighbors (KNN) is a basic and efficient non-parametric, supervised machine learning classifier which utilizes proximity to categorize or forecast the group to which a data item is allocated. The method operates by retaining all existing data points and categorizing new instances according to their similarity, generally assessed by distance metrics such as Euclidean, Manhattan, Minkowski, or Hamming distances [31]. The primary objective of KNN is to determine the nearest neighbors to a specified query point and subsequently allocate a class label. These distance functions facilitate the evaluation of proximate points, with the ultimate categorization established using a majority vote mechanism among the nearest neighbors. Each data point is categorized according to the predominant category among its closest neighbors, as established by the selected distance metric. The selection of K substantially influences prediction accuracy, lower values render the model susceptible to noise, whereas higher values elevate processing requirements. In datasets comprising two classes, researchers frequently select an odd value for K to prevent ties. Nonetheless, a limitation of KNN is that its processing performance may significantly diminish as the dataset size increases [30–33].

1.3. Back Propagation Neural Networks (BPNN)

Back Propagation Neural Networks (BPNNs) are the part of Artificial Neural Networks (ANNs) that employ the backpropagation method for training purposes. They are well acknowledged for their efficacy in deep learning models. A BPNN comprises a minimum of three layers of nodes: an input layer, one or more hidden layers, and an output layer. Every node, known as an artificial neuron or perceptron, that is interconnected by weighted linkages, which are modified throughout the training process [28,30,34,35]. Multilayer perceptrons employ the backpropagation method,

consisting of two phases: a forward pass, in which input data is processed through the network to produce an output, and a backward pass, during which the error (the differences between the anticipated and actually output) flows backward through the network to adjust the weights. This modification is executed utilizing optimization techniques, including gradient descent. Backpropagation Neural Networks (BPNNs) are advantageous due to their capacity to learn and represent complicated non-linear relationships between values[30,35]. Upon completion of training, algorithms may generate precise predictions when presented with novel data, rendering them exceptionally adaptable and versatile for real life usage. For a comprehensive elucidation of BPNN functionality, refer to Siddique et al. [29]. These models demonstrate proficiency in tasks necessitating pattern recognition and are trained utilizing features derived from diverse datasets, such images, numerical data, or text [36].

1.4. Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is a complex delicate machine learning technique derived from the gradient boosting framework. The fundamental idea is to enhance forecast performance by incrementally constructing an ensemble of decision trees, with each tree rectifying the weaknesses of its predecessors. In the training process, XGBoost initially establishes a rudimentary model (often a constant value) and subsequently computes the residuals or discrepancies between the predicted and actual values. In succeeding steps, a new decision tree is fitted to minimize these residuals using gradient descent, a method that modifies the model by pursuing the direction of maximal error reduction. New trees are incorporated into the ensemble to forecast the residuals, while the model adjusts the weights of misclassified examples to mitigate subsequent errors [37].

XGBoost integrates various distinctive optimizations. It employs regularization approaches (L1 and L2) to mitigate overfitting, hence assuring the model generalizes effectively to novel data [38]. The algorithm adeptly manages absent values and sparse datasets, autonomously identifying the optimal trajectory through the decision tree in the absence of data. XGBoost also executes tree pruning, ceasing tree growth when additional splits yield negligible enhancements, so improving both performance and efficiency [36]. XGBoost is esteemed for its scalability and quickness due to its capacity for parallel data processing, rendering it an optimal selection for managing extensive datasets and intricate issues. The aforementioned qualities have rendered XGBoost highly esteemed in both academic research and industrial applications, particularly in regression and classification tasks [36,37].

2. Materials and Methods

Study Design

A total of 94 intact male Spanish goats (58 healthy; 36 parasitized) (24 months old; 36-50 kg) were used in this experiment over multiple time points from August 2023 to October 2023, providing a robust dataset for analysis. This longitudinal approach allowed us to capture variations over time while adhering to ethical guidelines. All animal use protocols were approved by the Fort Valley State University (FVSU, Fort Valley, GA, USA) Agricultural and Laboratory Animal Care and Use Committee (ALACUC approval number F-T-01-2022). The study was carried out especially to ensure that regulations regarding animal welfare were followed by reducing any conditions that would cause the experimental goats' deaths, mainly caused by low PCV values. The goats were allowed to acquire a natural parasitic infection by grazing on grass pasture at the FVSU Agriculture Technology Center farm from March through September 2023. Healthy goats were monitored weekly for parasites and regularly dewormed using commercially available dewormers with veterinarian recommended doses. Several machines learning models, including Support Vector Machines (SVM), Back Propagation Neural Network, k-Nearest Neighbors (k-NN), XGBoost, and Keras deep learning models were trained and evaluated to determine their ability to classify goats based on their bioelectrical properties especially electrical resistance and electrical reactance. This study also aimed to assess the variations in bioelectrical impedance parameters such as electrical resistance (Rs) and

electrical reactance (X_c) between healthy and parasitized goats. A total of 1,540 observations points were collected from August 2023 to December 2023, consisting of 917 healthy goat observation points and 623 parasitized goats observation points from 94 goats throughout the experimental period. Bioelectrical impedance analysis (BIA) was employed to assess R_s and X_c , which reflect tissue hydration levels and cellular integrity.

3. Data Collection and Preprocessing

Data collection was conducted using an online cloud-based server specifically designed by the BIA device provider, which was preprocessed by dividing it into training and testing sets to assess model performance. For classification and regression tasks, features were scaled using standard techniques like normalization to ensure uniformity across the dataset [39]. This step was critical, especially for machine learning algorithms like Support Vector Machines (SVM) and K-Nearest Neighbors (KNN), which are sensitive to feature scaling [40,41]. The target variable 'y_test' was adjusted based on the problem type: classification models predicted categorical labels, while regression models predicted continuous values [42]. A 10-fold nested cross-validation approach was employed to reduce overfitting and evaluate model performance comprehensively (Varma & Simon, 2006). This method included an inner loop for hyperparameter optimization and an outer loop for model evaluation, thus providing more reliable performance estimates [43,44].

The statistical analyses were also conducted with PROC GLIMMIX in SAS to evaluate the impact of the condition (healthy versus parasitized) on R_s and X_c . The Poisson distribution with a logarithmic link function was employed to describe the response variables based on the characteristics of the data. Least squares mean (LS Means) were computed for each condition, and Tukey-Kramer corrections were applied to address multiple comparisons. Furthermore, PROC MEANS was employed to compute descriptive statistics, encompassing the mean, standard deviation, and standard error for each variable.

3.1. Model Development and Pipeline

The model development process comprised two primary components: classification and regression tasks, examined by several machine learning methods.

3.1.1. Classification Models Pipeline:

Backpropagation Neural Networks (BPNN), SVM, and KNN, were used for classification based analysis, each of which was evaluated using accuracy and AUC-ROC (Area Under the Receiver Operating Characteristic Curve) scores (Bishop, 2006). The 'classification_summary' function executed the process of predicting the test set outcomes and calculating the respective performance metrics [45]. The accuracy was computed using the 'accuracy_score' function from the scikit-learn library [45,46], which provides a direct comparison between the predicted and actual labels [47]. For models capable of generating probability estimates (e.g., SVM and BPNN), the AUC was calculated using the 'roc_auc_score' function, providing insight into how well the model distinguishes between different classes [48]. For models without probabilistic outputs (like KNN), the AUC score was marked as 'N/A' [42]. The results for each model were compiled into a panda 'DataFrame' to allow for the comparison of classification accuracy and AUC values [49].

3.1.2. Regression Models Pipeline:

The regression pipeline incorporated modified models for regression based analysis which includes BPNN, SVM, KNN, XGBoost, and a Keras-based Neural Network [37,50]. The 'regression_summary' function evaluated these models using two primary performance metrics: the R^2 score and Mean Squared Error (MSE). The R^2 score measured the proportion of variance in the test data explained by the model [51], while MSE quantified the average squared differences between the predicted and actual values, highlighting the overall prediction error [52]. This pipeline facilitated the comparison of model performance by storing these results in a 'pandas DataFrame', providing a

clear view of the most effective regression algorithms for the given dataset [49]. The integration of a Keras neural network was particularly useful in exploring the capabilities of deep learning approaches in solving regression tasks [53].

3.1.3. Nested Cross-Validation

A 10-fold nested cross-validation method was employed to enhance model resilience and mitigate the possibility of overfitting [43,54]. In the outer loop, the dataset comprising 1540 observation points was partitioned into 10 equal folds, with each fold utilized as the test group once, while the remaining nine folds or groups were employed for training. Hyperparameter optimization for the inner loop was conducted using an additional 10-fold cross-validation within the training set, ensuring the model's parameters were refined without incorporating test data into the training phase [43]. This nested technique yielded more dependable performance estimates by considering both model selection and model evaluation [42,44].

The layered cross-validation technique facilitated thorough optimization of model hyperparameters, especially for models like SVM and XGBoost, which necessitate meticulous parameter selection for peak performance [37,55]. The inner cross-validation loop guaranteed that model selection relied on optimal parameters derived from training data, without affecting the model's performance assessment in the outer loop [54]. This validation method was essential for both classification and regression processes, guaranteeing that the final performance measures appropriately represented the model's generalization capacity [44]. The performance measures from all outer folds were summed to yield a reliable assessment of the model's efficacy throughout the complete dataset [43]. This complex validation method produced more dependable models, especially in instances where the used dataset demonstrated variability [56]. It also improves the confidence in model predictions for both classification and regression tasks, rendering it a significant element of our investigation [44,50,56].

4. Results and Discussion

The PROC GLIMMIX method revealed a significant impact of condition (healthy vs. parasitized) on electrical resistance (Rs) ($F = 635.36$, $P < 0.0001$). Parasitized goats demonstrated a significantly higher least squares mean electrical resistance (5.5321 ± 0.002520 SE) in comparison to healthy goats (5.4483 ± 0.002166 SE). The difference between the two categories was significantly different (Tukey-adjusted $P < 0.0001$), with parasitized goats exhibiting an increase of 0.08377 units in electrical resistance. The PROC MEANS approach yielded descriptive statistics, revealing that the mean Rs for healthy goats was 232.37 ± 5.77 SE, but parasitized goats demonstrated an average mean Rs of 252.68 ± 7.31 SE, hence corroborating the enhanced electrical resistance in parasitized animals.

This study's results indicate that parasite infections markedly affect the bioelectrical impedance parameters of goats, especially electrical resistance (Rs). Parasitized goats exhibited a significant elevation in Rs relative to healthy goats, potentially attributed to blood loss, dehydration, and tissue changes induced due to the GIN parasite infection [57]. The parasite feeds on the host's blood, resulting in anemia and lower fluid volume, hence elevating tissue electrical resistance when electrical current encounters increased opposition in drier, less hydrated tissues [1,58,59].

A notable difference in electrical reactance (Xc) was also observed between healthy and parasitized goats ($F = 11.12$, $P = 0.0009$). Parasitized goats had greater least squares mean electrical reactance (3.9628 ± 0.005524 SE) than healthy goats (3.9388 ± 0.004608 SE), although the disparity was less pronounced than that observed for electrical resistance. The Tukey-adjusted comparison indicated statistical significance ($P = 0.0009$), with a mean difference of 0.02399 units in electrical reactance. Descriptive statistics using PROC MEANS indicated mean Xc values of 51.36 ± 4.26 SE for healthy goats and 52.60 ± 5.23 SE for parasitized goats. Although the difference in electrical reactance was statistically significant, the size of the change was lesser than that of electrical resistance, suggesting a less prominent effect.

The rise in electrical reactance (Xc) in parasitized goats was statistically significant, although the impact was lower to that of electrical resistance (Table 1). Electrical reactance indicates cell membrane

integrity and fluid distribution, and the slight rise in X_c implies that parasite infections exert a limited influence on the capacitive characteristics of tissues [60]. This may result from tissue injury and cell membrane disruption, which influence the storage of electrical current in the tissues, however the impact is less significant than electrical resistance [61–64]

Table 1. Comparison of bioelectrical impedance parameters (Electrical resistance and Electrical reactance) between healthy and parasitized goats, measured using bioelectrical impedance analysis (BIA).

Parameter	Condition	
	Healthy	Parasitized
Electrical resistance (R_s)	232.37 ± 5.07 ^b	252.67 ± 7.32 ^a
Electrical reactance (X_c)	51.36 ± 4.26 ^y	52.60 ± 5.24 ^x

Superscript on mean value for condition is significantly different ($p < 0.009$) for electrical resistance and electrical reactance between healthy and parasitized goats.

Table 2 summarizes the classification performance of the models, presenting the Accuracy, Precision, Recall, and F1-Score for each model. Among the models, the SVM exhibited superior performance, achieving an accuracy of 95%, a precision of 93%, and an F1-Score of 96%, demonstrating its robust capacity to differentiate between healthy and parasitized goats. XGBoost achieved an accuracy of 94% and an F1-Score of 95%, indicating the efficacy of ensemble approaches in managing intricate datasets [37,50]. The BPNN model demonstrated a high level of performance, with an accuracy of 92% and an F1-Score of 94%, positioning it as a competitive alternative for classification tasks [65,66]. Keras DL attained an accuracy of 91% and an F1-Score of 93%, indicating that deep learning models can perform effectively, although they were unable to exceed SVM or XGBoost [67,68]. Ultimately, K-NN attained the lowest accuracy of 90%, suggesting it underperformed relative to more advanced models such as SVM and XGBoost, perhaps because to its simplicity and susceptibility to data noise and non-linearity in data [67].

Table 2. Comparison of different classification performance matrices for classification of goat health condition (healthy Vs parasitized).

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	95	93	94	96
BPNN	92	91	92	94
K-NN	90	88	89	92
XGBoost	94	92	93	95
Keras DL	91	90	91	93

This Table presents the results for the different model performance matrices for classifying goat health to detect parasitized vs healthy goats. Accuracy, Precision, Recall, and F1 Score are performance metrics represented in percentage that assess the model's classification efficiency. The F1 score is the harmonic mean of Precision and Recall, providing a balanced measure of the model's ability to classify data correctly.

The robust efficacy of SVM and XGBoost in classification tasks can be linked to several factors. Support Vector Machine (SVM) operates by identifying the ideal hyperplane that maximizes the margin between classes, rendering it particularly successful for linearly separable datasets [42,50,55,69]. Considering that bioelectrical impedance measurements probably demonstrate unique patterns between healthy and parasitized goats, the capability of SVM to establish a definitive separation in the feature space enables it to attain high accuracy. Furthermore, SVM excels with high-dimensional data, potentially demonstrating its higher efficacy compared to simpler models such as K-NN [33,42].

XGBoost demonstrated excellent performance, which is anticipated due to its capacity to manage intricate, non-linear interactions via boosting. XGBoost captures complex patterns in the dataset by systematically rectifying errors from prior iterations [37,50]. This is especially advantageous in biological datasets because nuanced variations in characteristics can significantly influence classification [42,70]. The BPNN and Keras deep learning models exhibited marginally reduced performance compared to SVM and XGBoost, although they still demonstrated competitive outcomes [65]. Neural networks probably encapsulate intricate, non-linear relationships within the data; yet their efficacy may be affected by the selection of hyperparameters or the dimensions of the neural network. The K-NN method, due to its reliance on proximity-based judgments, had difficulties managing the dataset's complexity, maybe accounting for its lower accuracy relative to other models [51,68].

Table 3 shows the efficacy of the models in forecasting the extent of parasitism, as indicated by the R-squared (R^2) value and Mean Squared Error (MSE). The BPNN model demonstrated superior performance, achieving a R^2 value of 99.9% and a minimal MSE of 1.25e-04, signifying its exceptional predictive accuracy about the health status of goats [65]. SVR demonstrated strong performance, achieving a R^2 value of 96.9% and a minimal MSE of 7.69e-03, indicating its high reliability as a regression model [37,50]. XGBoost and Keras DL attained R^2 values of 89.2% and 88%, respectively, with moderate MSE values, suggesting that although these models exhibited better performance, they lacked the precision of BPNN or SVR. K-NN regression exhibited the poorest performance, with a R^2 value of 83% and a higher MSE of 3.30e-02, indicating its worse ability to describe the intricate correlations between bioelectrical data and goat health problems relative to the other models [33].

Table 3. Comparison of different regression model performance matrices for classification of goat health condition (Healthy Vs Parasitized).

Models	R square Value	MSE
SVR	96.9	7.69e-03
BPNN	99.9	1.25e-04
KNN	83.0	3.30e-02

XGBoost	89.2	0.025
Keras DL	88	0.027

The table presents a comparison of various models used for regression analysis, with their corresponding R-squared values and Mean Squared Error (MSE). These results provide insights into the performance of different algorithms, highlighting the superior accuracy of the BPNN model in this specific context.

The BPNN regression model was the most successful model in regression tasks, obtaining a nearly perfect R^2 value of 99.9%. This suggests that the neural network model was able to accurately anticipate the degree of parasitism by capturing nearly all of the variance in the dataset [65,68]. The success of BPNN in regression can be attributed to its capacity to learn complex, non-linear patterns, which is particularly advantageous in biological data that involve interactions between various physiological parameters. BPNN outperformed other models in regression tasks because neural networks are well-suited for encoding these intricate relationships [71].

Additionally, SVR demonstrated better performance, attaining a R^2 value of 96.9%. SVR operates similarly to SVM, but it is designed to identify a function that deviates from the true data points by a small margin for continuous data [72]. The efficacy of SVR in predicting parasitism severity may have been influenced by its capacity to manage outliers and noisy data. XGBoost and Keras DL also demonstrated strong predictive potential; however, they were unable to achieve the same level of precision as BPNN or SVR. This could be attributed to the hyperparameter tuning limitations or the complexity of the data [37,51]. However, K-NN regression encountered difficulty with this task, most likely due to its dependence on local averaging, which may not adequately capture the global trends apparent in the dataset, as opposed to more advanced models such as BPNN and SVR [32,68].

5. Conclusions

Bio-electrical impedance analysis approach has the potential to revolutionize parasitism detection and livestock management practices. Traditional methods like FEC and Hematocrit Analysis, while effective, require substantial labor, time, and expertise, making them less practical for large-scale or resource-constrained operations. In contrast, BIA, combined with machine learning, can provide real-time diagnostic information without the need for invasive procedures or specialized personnel. This could significantly enhance the ability of farmers and veterinarians to monitor the health of their herds, enabling earlier detection and treatment of parasitic infections, reducing production losses, and improving animal welfare.

This study shows the importance of bioelectrical impedance analysis (BIA) in goat parasitism diagnosis. Machine learning models effectively classified goats as healthy or parasitized based on bioelectrical impedance. Support Vector Machine (SVM) and Backpropagation Neural Networks (BPNN) were the most successful models, with BPNN obtaining near-perfect accuracy in predicting parasitism (R^2 value of 99.9%). BPNN's capacity to capture complicated, non-linear biological data correlations is vital for dealing with physiological fluctuations in parasitized animals, explaining its excellent performance. Parasitic infections alter goat tissue electrical characteristics, as seen by differences in electrical resistance (R_s) and electrical reactance (X_c) between healthy and parasitized goats. Due to dehydration and blood loss from parasitic diseases like *Haemonchus contortus*, parasitized goats had increased electrical resistance (R_s). These illnesses cause anemia, which reduces blood volume and tissue fluid retention, raising electrical current electrical resistance. This supports prior studies indicating that parasite infections in small ruminants alter hydration and tissue composition, which bioelectrical impedance may detect. High classification accuracy of machine learning models like SVM and XGBoost shows their robustness in high-dimensional, complicated datasets like bioelectrical impedance measurements. These models are ideal for real-time diagnostics because they can detect minor differences between healthy and parasitized animals.

Future Research

Future research can build on this study in multiple ways. First, bioelectrical impedance analysis (BIA) can be used to diagnose more animal health issues. BIA can detect tissue changes caused by metabolic abnormalities, dietary deficits, and chronic illnesses. Studying BIA in these circumstances could yield useful insights and non-invasive ways for early diagnosis of livestock (including ruminants and small ruminants) health concerns. Further research into deep learning models may improve livestock examinations machine learning prediction. Test more advanced neural networks like CNNs or RNNs to determine whether they improve the present models. When paired with larger datasets or time-series data from repeated BIA tests, these models may better capture temporal trends and long-term dependencies. Researchers could construct algorithms that detect and forecast parasitism by tracking goats and collecting BIA data at different phases of infection. Early intervention and improved herd health and productivity may result from more effective treatment regimens. These methods could also be applied to sheep, and cattle to test the generalizability of BIA and machine learning models across farming systems. Understanding how parasitism or other health issues affect each species' bioelectrical characteristics can help build species-specific diagnostic tools. Finally, future research should also examine the economic benefits of BIA and machine learning for health issues identification. Early infection prevention can save treatment costs, productivity losses, and animal welfare, improving farm profitability and sustainability.

Author Contributions: A.S: as the first author and corresponding author of this original research, conceptualized the research methodology in consultations with the entire research team participating in this manuscript as coauthors. S.S.P., A.S, P.B., D.B., P.G., A.S, and T.E. completed the data collection, data processing, analyses, preparation and revision of manuscript. DS-I, J.A.V.W. and E.R.M participated in the preparation and revision of the manuscript. T.H.T. G.K., and A.K.M. provided technical help on assessment and ground truthing of the BIA data, study design and modeling, and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by USDA-ARS.

Data Availability Statement: Data can be provided upon request to Corresponding Author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Arsenopoulos KV, Fthenakis GC, Katsarou EI, Papadopoulos E. Haemonchosis: A challenging parasitic infection of sheep and goats. *Animals*. 2021;11(2):363.
2. Bunyan A. Investigations into the population structure of the ovine parasitic nematode *Haemonchus contortus*: A comparison of isolates from differing climatic and geographical regions of New South Wales [Internet]. Charles Sturt University; 2021. <https://researchoutput.csu.edu.au/en/publications/investigations-into-the-population-structure-of-the-ovine-parasit> [Accessed 2024-10-10].
3. Casey SJ. *Haemonchus contortus* infections in alpacas and sheep [Doctoral dissertation]. Virginia Tech; 2014.
4. Teddleton HG. Effect of *Haemonchus contortus* excretory/secretory protein on differences in host neutrophil migration; 2024.
5. Clark DAS. *Haemonchus contortus* and hookworms—parallels in vaccine development [Doctoral dissertation]. University of Glasgow; 2006.

6. Glaji YA, Mani AU, Bukar MM, Igbokwe IO. Reliability of FAMACHA© chart for the evaluation of anaemia in goats in and around Maiduguri. Sokoto J Vet Sci. 2014;12(3):9-14.
7. Kaplan RM, Burke JM, Terrill TH, Miller JE, Getz WR, Mobini S, et al. Validation of the FAMACHA© eye color chart for detecting clinical anemia in sheep and goats on farms in the southern United States. Vet Parasitol. 2004;123(1-2):105-20.
8. Cain JL, Slusarewicz P, Rutledge MH, McVey MR, Wielgus KM, Zynda HM, et al. Diagnostic performance of McMaster, Wisconsin, and automated egg counting techniques for enumeration of equine strongyle eggs in fecal samples. Vet Parasitol. 2020;284:109199.
9. Verocai GG, Chaudhry UN, Lejeune M. Diagnostic methods for detecting internal parasites of livestock. Vet Clin Food Anim Pract. 2020;36(1):125-43.
10. Ljungström S, Melville L, Skuce PJ, Höglund J. Comparison of four diagnostic methods for detection and relative quantification of *Haemonchus contortus* eggs in feces samples. Front Vet Sci. 2018;4:239.
11. Levecke B, Kaplan RM, Thamsborg SM, Torgerson PR, Vercruysse J, Dobson RJ. How to improve the standardization and the diagnostic performance of the fecal egg count reduction test?. Vet Parasitol. 2018;253:71-8.
12. Bosco A. The coprological diagnosis of gastrointestinal nematode infections in small ruminants [Doctoral dissertation]. Napoli: Università degli Studi di Napoli Federico II; 2014. <https://scholar.google.com/scholar>
13. Demelash K, Abebaw M, Negash A, Alene B, Zemene M, Tilahun M. A review on diagnostic techniques in veterinary helminthology. Nat Sci. 2016;14(7):109-18.
14. Bentounsi B, Attir B, Meradi S, Cabaret J. Repeated treatment faecal egg counts to identify gastrointestinal nematode resistance in a context of low-level infection of sheep on farms in eastern Algeria. Vet Parasitol. 2007;144(1-2):104-10.
15. Rinaldi L, Veneziano V, Morgoglione ME, Pennacchio S, Santaniello M, Schioppi M, et al. Is gastrointestinal strongyle faecal egg count influenced by hour of sample collection and worm burden in goats?. Vet Parasitol. 2009;163(1-2):81-6.
16. Ngere L, Burke JM, Morgan JLM, Miller JE, Notter DR. Genetic parameters for fecal egg counts and their relationship with body weights in Katahdin lambs. J Anim Sci. 2018;96(5):1590-9.
17. Valentinuzzi ME, Morucci JP, Felice CJ. Bioelectrical impedance techniques in medicine part II: monitoring of physiological events by impedance. Crit Rev Biomed Eng. 1996;24(4-6).
18. Kushner RF. Bioelectrical impedance analysis: a review of principles and applications. J Am Coll Nutr. 1992;11(2):199-209.
19. Davydov DM, Boev A, Gorbunov S. Making the choice between bioelectrical impedance measures for body hydration status assessment. Sci Rep. 2021;11(1):7685.
20. Gadir G, Gunay K. Measurement of electrical conductivity of biologically active points. Endless Light Sci. 2023;(May):281-6.
21. Marra M, Sammarco R, De Lorenzo A, Iellamo F, Siervo M, Pietrobelli A, et al. Assessment of body composition in health and disease using bioelectrical impedance analysis (BIA) and dual-energy X-ray absorptiometry (DXA): a critical overview. Contrast Media Mol Imaging. 2019;2019(1):3548284.
22. Jackson AA, Johnson M, Durkin K, Wootton S. Body composition assessment in nutrition research: value of BIA technology. Eur J Clin Nutr. 2013;67(1)
23. Vapnik, V., Golowich, S., & Smola, A. (1996). Support vector method for function approximation, regression estimation and signal processing. *Advances in neural information processing systems*, 9.

24. Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T., & Vapnik, V. (2000). Feature selection for SVMs. *Advances in neural information processing systems*, 13.
25. Winston, P.H. (2024). Artificial Intelligence Course in MIT Open Course Ware.
26. Cortes C. Support-Vector Networks. Machine Learning. 1995.
27. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. arXiv 2012. arXiv preprint arXiv:1207.0580. 2012.
28. Siddique A, Herron CB, Wu B, Melendrez KS, Sabillon LJ, Garner LJ, Durstock M, Sanz-Saez A, Morey A. Development of Predictive Classification Models and Extraction of Signature Wavelengths for the Identification of Spoilage in Chicken Breast Fillets During Storage Using Near Infrared Spectroscopy. *Food and Bioprocess Technology*. 2024 Jul 13:1-9.
29. Siddique, A., Shirzaei, S., Smith, A. E., Valenta, J., Garner, L. J., & Morey, A. (2021). Acceptability of artificial intelligence in poultry processing and classification efficiencies of different classification models in the categorisation of breast fillet myopathies. *Frontiers in Physiology*, 12, 712649.
30. Siddique A. IMPLEMENTING BIG DATA ANALYTICS APPROACHES TO IMPROVE FOOD QUALITY AND MINIMIZE FOOD WASTE AND LOSS.
31. Siddique A, Herron CB, Valenta J, Garner LJ, Gupta A, Sawyer JT, Morey A. Classification and feature extraction using supervised and unsupervised machine learning approach for broiler woody breast myopathy detection. *Foods*. 2022 Oct 20;11(20):3270.
32. Halder RK, Uddin MN, Uddin MA, Aryal S, Khraisat A. Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. *Journal of Big Data*. 2024 Aug 11;11(1):113.
33. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat*. 1992;46(3):175-85. <https://doi.org/10.1080/00031305.1992.10475879>
34. Vasconcelos GA, Francisco MB, da Costa LR, Ribeiro Junior RF, Melo MD. Prediction of surface roughness in duplex stainless steel face milling using artificial neural network. *The International Journal of Advanced Manufacturing Technology*. 2024 Jun 8:1-8.
35. Pratap B, Bansal S. Optimizing Artificial Neural-Network Using Genetic Algorithm. *Bio-Inspired Optimization for Medical Data Mining*. 2024 Jul 19:269-88.
36. Padhy R, Dash SK, Khandual A, Mishra J. Image classification in artificial neural network using fractal dimension. *International Journal of Information Technology*. 2023 Aug;15(6):3003-13.
37. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. *Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min*. 2016:785-94. <https://doi.org/10.1145/2939672.2939785>
38. Utkarsh, Jain PK. Predicting bentonite swelling pressure: optimized XGBoost versus neural networks. *Scientific Reports*. 2024 Jul 26;14(1):17533.
39. Cao Y, He Y, Bai H. Feature scaling optimization in machine learning. *IEEE Access*. 2020;8:112154-65.
40. Roy S, Mukherjee A, Biswas A. A comprehensive study of scaling in machine learning models. *Int J Mod Trends Sci Technol*. 2021;7(2):11-7.
41. Zhang S, Yao L, Sun A, Tay Y. Deep learning based recommender system: A survey and new perspectives. *ACM computing surveys (CSUR)*. 2019 Feb 25;52(1):1-38.
42. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning. New York: Springer; 2009. <https://doi.org/10.1007/978-0-387-21606-5>
43. Cawley GC, Talbot NL. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Mach Learn Res*. 2010;11:2079-107.

44. Varoquaux G, Raamana PR, Engemann DA, Hoyos-Idrobo A, Schwartz Y, Thirion B. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. *Neuroimage*. 2017;145:166-79.
45. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res*. 2011;12:2825-30.
46. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. *Science*. 2015;349(6245):255-60.
47. Kumar D, Das S, Kumar D. Ensemble learning techniques: An overview. *Int J Mach Learn Netw Collab Eng*. 2021;1(1):1-12.
48. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit*. 1997;30(7):1145-59.
49. McKinney W. Data structures for statistical computing in Python. In: *Proc 9th Python Sci Conf*. 2010;445:51-6.
50. Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Stat*. 2001;29(5):1189-232. <https://doi.org/10.1214/aos/1013203451>
51. Goodfellow I, Bengio Y, Courville A. *Deep learning*. MIT Press; 2016.
52. Seber GAF, Lee AJ. *Linear regression analysis*. John Wiley & Sons; 2012.
53. Shome A, Mukherjee G, Chatterjee A, Tudu B. Study of Different Regression Methods, Models and Application in Deep Learning Paradigm. In *Deep Learning Concepts in Operations Research 2024 Aug 30* (pp. 130-152). Auerbach Publications.
54. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*. 2006;7(1):1-8.
55. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273-97. <https://doi.org/10.1007/BF00994018>
56. Kuhn M, Johnson K. *Applied predictive modeling*. Springer; 2013.
57. Aburto-Corona JA, Calleja-Núñez JJ, Moncada-Jiménez J, de Paz JA. The effect of passive dehydration on phase angle and body composition: A bioelectrical impedance analysis. *Nutrients*. 2024;16(14):2202.
58. Hioka A, Akazawa N, Okawa N, Nagahiro S. Extracellular water-to-total body water ratio is an essential confounding factor in bioelectrical impedance analysis for sarcopenia diagnosis in women. *Eur Geriatr Med*. 2022;13(4):789-94.
59. Hoste H, Torres-Acosta JFJ, Quijada J, Chan-Perez I, Dakheel MM, Kommuru DS, et al. Interactions between nutrition and infections with *Haemonchus contortus* and related gastrointestinal nematodes in small ruminants. *Adv Parasitol*. 2016;93:239-51.
60. Ward LC, Brantlov S. Bioimpedance basics and phase angle fundamentals. *Rev Endocr Metab Disord*. 2023;24(3):381-91.
61. Brito DRB, Júnior LMC, Garcia JL, Chaves DP, Júnior JAAC, Conceição WLF, de Brito AVM. Clinical parameters of goats infected with gastrointestinal nematodes and treated with condensed tannin. *Semina Ciênc Agrár*. 2020;41(2):517-30.
62. Shim G, Breinyn IB, Martínez-Calvo A, Rao S, Cohen DJ. Bioelectric stimulation controls tissue shape and size. *Nat Commun*. 2024;15(1):2938.
63. Moro AB, Galvani DB, Montanholi YR, Bertemes-Filho P, Venturini RS, Martins AA, da Silva LP, Pires CC. Assessing the composition of the soft tissue in lamb carcasses with bioimpedance and accessory measures. *Meat science*. 2020 Nov 1;169:108192.

64. Moro C, Stromberga Z, Moreland A. Enhancing teaching in biomedical, health and exercise science with real-time physiological visualisations. *Biomedical Visualisation: Volume 8*. 2020:1-1.
65. Hecht-Nielsen R. Theory of the backpropagation neural network. *Proc Int Joint Conf Neural Netw*. 1988:593-605. <https://doi.org/10.1109/IJCNN.1988.118638>
66. Bishop CM. *Neural Networks for Pattern Recognition*. Oxford University Press; 1995.
67. Gulli A, Kapoor A, Pal S. *Deep learning with TensorFlow 2 and Keras: regression, ConvNets, GANs, RNNs, NLP, and more with the Keras API*. Packt Publishing Ltd; 2019.
68. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-44. <https://doi.org/10.1038/nature14539>
69. Scholkopf B, Smola AJ. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT Press; 2002.
70. Caruana R, Niculescu-Mizil A. An empirical comparison of supervised learning algorithms. *Proc 23rd Int Conf Mach Learn (ICML)*. 2006:161-8. <https://doi.org/10.1145/1143844.1143865>
71. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986;323(6088):533-6.
72. Smola AJ, Schölkopf B. A tutorial on support vector regression. *Stat Comput*. 2004;14(3):199-222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.