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

Kafkas Univ Vet Fak Derg
xx (x): xx, 2025
DOI: 10.9775/kvfd.2025.34031

RESEARCH ARTICLE

Hybrid Ensemble Model for Lactation Milk Yield Prediction of Holstein Cows

Derviş TOPUZ [1 (*)] Selçuk TEKGÖZ [2]

[1] Niğde Ömer Halisdemir University, Niğde Zübeyde Hanım Vocational School of Health Services, Department of Health Services Science, TR-51240 Niğde-TÜRKİYE

[2] Niğde Ömer Halisdemir University, Graduate School of Natural and Applied Sciences, Department of Interdisciplinary Disaster Management, TR-51240 Niğde-TÜRKİYE

(*) **Corresponding author:**
Derviş Topuz
Phone: +90 388 221 2806
Cellular phone: +90 506 881 3838
E-mail: topuz@ohu.edu.tr

How to cite this article?

Topuz D, Tekgöz S: Hybrid ensemble model for lactation milk yield prediction of Holstein cows. *Kafkas Univ Vet Fak Derg*, 2025 (Article in Press).
DOI: 10.9775/kvfd.2025.34031

Article ID: KVFD-2025-34031
Received: 07.03.2025
Accepted: 13.08.2025
Published Online: 19.08.2025

Abstract

Machine learning (ML) algorithms are widely employed across various domains to identify patterns and relationships in large datasets, and to perform tasks such as prediction and classification. This study investigates the use of machine learning techniques to predict lactation milk yield in Holstein dairy cows within the field of veterinary sciences. The dataset comprises records from 128 cows, with lactation milk yield categorized into three classes low, medium, and high based on threshold values determined by expert opinion. The independent variables include Age (in days), Days in Milk (DIM), Service Period (in days), Calving Date, and Parity. To reduce the dimensionality of the dataset, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were applied. The performance of nine classification algorithms was evaluated on both the original and reduced datasets using 10-fold cross-validation and bootstrap resampling methods. Due to class imbalance in the data, the weighted F1-score was used as the primary performance metric instead of accuracy. Among the original models, the highest weighted F1-scores were achieved by Decision Tree (DT), Gradient Boosting Machine (GBM), and Extreme Gradient Boosting (XGBoost), with scores of 0.47, 0.53, and 0.51, respectively. A hybrid ensemble model developed by combining these top-performing algorithms demonstrated superior performance, yielding a weighted F1-score of 1.00, an accuracy of 1.00, and an ROC-AUC of 1.00. These findings suggest that hybrid ensemble models can provide more effective and robust solutions in veterinary applications and similar research fields.

Keywords: Machine Learning, Decision tree, Gradient boosting machine, Xgboost, Milk yield, Holstein**,** F1-score, AUC, Hybrid model

INTRODUCTION

Livestock production is vital for global food security and demands ongoing research to boost productivity. Effective herd management enhances animal welfare, health, and profitability, which are key to sustainable milk production [1,2]. Machine learning (ML), a branch of artificial intelligence (AI), offers advanced tools to detect complex patterns, select relevant variables, and generate accurate predictions in this context [3]. ML and AI are widely used in fields such as veterinary science, enabling improved decision-making through big data and IoT technologies [4-6]. Recent advances have accelerated AI applications in livestock, including disease detection and biometric animal monitoring, enhancing real-time herd management [4,7-12].

Many studies apply ML to predict milk yield, but results often lack reliability under real farm conditions due to milk yield's multifactorial complexity [13,14]. Traditional accuracy metrics can mislead in imbalanced datasets; therefore, weighted F1-score and ROC-AUC are preferred for performance evaluation [15]. ML has shown promise in disease detection, reproductive performance prediction, and resource optimization in dairy farming [16-20].

This study evaluated nine ML algorithms to classify Holstein cows' lactation milk yield using a dataset of 128 records with five features. Dimensionality reduction via PCA and LDA was applied. Models were assessed using 10-fold cross-validation and 50 bootstrap samples, focusing on weighted F1-score and ROC-AUC due to class imbalance. A hybrid ensemble model combining top algorithms was developed to improve prediction accuracy and robustness. The following sections detail the data, methods, results, and the enhanced performance of the hybrid model.

Material and Methods

Ethical Statement

This study does not require ethical permission.

Data and Preprocessing

The study dataset comprised 128 lactation records from Holstein cows, each with five features: age (days), days in milk (DIM), service period (days), calving date, and parity. The target variable, total lactation milk yield (kg), was classified into three categories (low, medium, high) based on expert thresholds, forming a multi-class classification task. Numerical features were normalized, categorical data numerically encoded, and outliers were removed using the Local Outlier Factor (LOF) method [21] excluding nine samples. A robust scaler minimized the effect of extreme values. Descriptive statistics and correlation analysis were conducted to examine feature relationships and multicollinearity. Class imbalance was confirmed, necessitating weighted metrics for model evaluation.

Dimensionality reduction using Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) was applied to reduce redundancy and noise: PCA reduced features from five to three components (97% variance retained), while LDA projected data onto two discriminant axes maximizing class separability. These reduced datasets were used to train classifiers, but as no significant performance gain was observed, hyperparameter tuning was conducted on models using the full feature set (128x5).

Machine Learning Algorithms

ML algorithms identify patterns in large datasets, enabling accurate predictions on unseen data [22]. Successful ML application depends on choosing suitable algorithms and evaluation metrics [23,24]. *Fig.1* illustrates a typical ML approach.

Common ML tasks include classification and regression, which require different algorithms. In this study, nine supervised classifiers were implemented using Python 3.9.10 and Scikit-learn libraries.

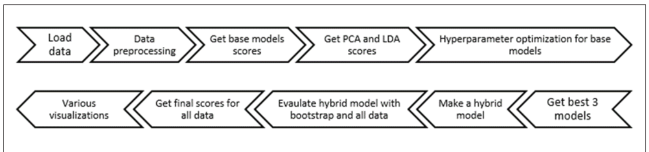

Fig 1. Workflow diagram

Decision Tree (DT): Tree-structured models split data by features and provide interpretable "if-then" rules. Pruning and tuning prevent overfitting [25,26].

Gradient Boosting Machine (GBM): An ensemble method building trees sequentially to correct prior errors; sensitive to hyperparameters like learning rate and tree depth [26].

Extreme Gradient Boosting (XGBoost): An optimized GBM variant with regularization and faster training, widely recognized for superior performance [27], which has demonstrated strong performance in agricultural datasets [28], also exhibits high tolerance to multicollinearity and missing data scenarios [29].

Random Forest (RF): Ensemble of decision trees trained on bootstrap samples with random feature subsets, robust to overfitting and nonlinearities [28-32].

K-Nearest Neighbors (KNN): Instance-based method classifying by majority vote of nearest neighbors using distance metrics; effective but computationally expensive in high dimensions [33].

Hybrid Model: A voting ensemble combining DT, GBM, and XGBoost leveraged their complementary strengths to improve classification of milk yield (low, medium, high). Ensemble consensus reduces misclassification risk and enhances generalization.

Performance Evaluation Metrics and Validation

Due to class imbalance, traditional accuracy can be misleading [11,34,35]. Thus, multiple metrics were used:

Accuracy: Ratio of correct predictions but less informative for imbalanced data *(Eq. 1)*.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{FP+FN+TP+TN}} \quad (1)$$

Here, *FP* denotes false positives and *FN* denotes false negatives.

Precision (P): Correct positive predictions among all positive predictions *(Eq. 2)*.

$$\text{Precision (P)} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

Recall (R): Correct positive predictions among all actual positives *(Eq. 3)*.

$$\text{Recall (R)} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

F1-score: Harmonic mean of precision and recall, balancing sensitivity and specificity, prioritized here with class-weighting to handle imbalance *(Eq. 4)* [36,37].

$$\text{F}_1 = 2 \times \frac{\text{P}\times\text{R}}{\text{P+R}} \quad (4)$$

ROC-AUC: Threshold-independent measure of discrimination ability, calculated as weighted average over all class pairs (OvO approach) [38].

Validation Techniques

10-Fold Cross-Validation: Dataset split into ten folds; each fold used once as test data to provide unbiased performance estimates (10% test, 90% train per fold) [39].

Bootstrap Sampling: To rigorously assess the statistical reliability, stability, and generalizability of model performance, a systematic bootstrap resampling strategy was employed in this study. This technique entails the generation of multiple resampled datasets by randomly drawing observations from the original dataset with replacement, thereby facilitating an evaluation of model behavior under varied sampling scenarios. Specifically, 50 independent bootstrap samples, each consisting of 10 randomly selected observations, were generated. These samples were subsequently used to investigate the variability in model predictions and to estimate performance stability and robustness [40]. Remarkably, the results demonstrated consistently perfect performance, with both accuracy and weighted F1-scores achieving 1.00 across all resampled datasets. These findings provide strong empirical evidence supporting the robustness, reliability, and invariance of the proposed models across varying data subsets. Furthermore, the bootstrap approach functioned as a powerful statistical tool for deriving more reliable estimates of model variance and predictive error, particularly in contexts characterized by limited data availability and potential shifts in data distribution. Overall, the methodology reinforces confidence in the models' generalization capability beyond the original training data, ensuring dependable performance in real-world applications.

Hyperparameter Tuning and Model Optimization

Hyperparameters external model settings were optimized via grid search combined with 10-fold CV to maximize weighted F1-score. This ensured balanced performance across classes and avoided overfitting. Optimization was performed for DT, GBM, and XGBoost.

Hyperparameters were chosen considering model complexity, overfitting risk, and class differentiation. After tuning, each algorithm was retrained on the full dataset with optimal hyperparameters. This improved both individual model performance and the hybrid model's overall effectiveness. Results from the original and dimensionally reduced datasets are presented, comparing individual models and the hybrid ensemble. Evaluation focuses on metrics, especially the weighted F1-score, with comprehensive discussion of findings.

Results

The creation of effective AI- and ML-based prediction systems requires a robust data pipeline, including: (1) data collection, (2) transformation into suitable formats, (3) secure storage, (4) analytical modeling and (5) presenting interpretable results. In this study, these steps were systematically applied to predict lactation milk yield. Nine classification algorithms were tested, and the effects of dimensionality reduction and model choice on performance were compared.

Step 1 - Data Collection: Relevant data were systematically gathered for analysis.

Step 2 - Data transformation: All features were standardized using the Robust Scaler *(Eq. 5)* and nine outliers were removed with the LOF algorithm [21].

Robust Scaler $$X_{new} = \frac{X - X_{median}}{IQR} \quad (5)$$

Step 3,4 - Data storage and data analysis: Basic statistics and correlations were examined to understand dataset

Table 1. *Calculated Statistical Values for the Categories of the Dependent Variable (n=128)*

Class	Number	Min.	Average	Median	Max.	Lower Bound	Upper Bound	Decision Boundaries
Low Milk Yield (0)	39	6208	7067.95	7203	7561	6926.22	7209.67	7561.2
Medium Milk Yield (1)	63	7563	8269.02	8296	8903	8168.42	8369.62	8915.6
High Milk Yield (2)	26	8924	9401.65	9409.5	9966	9289.31	9514	

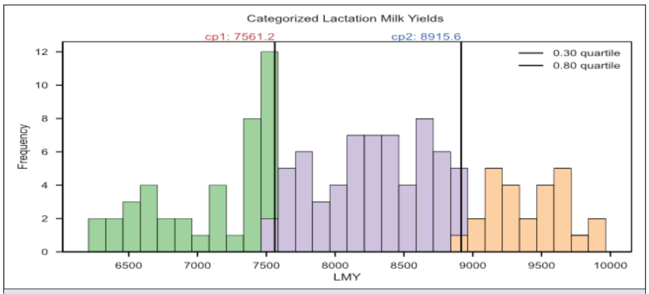

Fig 2. Classification of lactation milk yield values for 128 Holstein cows (Classification Model)

structure. A moderate negative correlation (-0.7) was found between the Milking Days Count (MDC) and Calving Date (CD), suggesting that cows with later calving dates tended to have fewer milking days. Based on expert opinion, the continuous dependent variable, lactation milk yield, was categorized into three classes (low, medium, high) using the 30th and 80th percentiles (7.561,2 kg; 8.915,6 kg). Class imbalance was addressed by assigning higher weights to minority classes [30]. Class distribution is shown in *Table 1*, with a visual representation in *Fig. 2*.

Class imbalance occurs when some classes have far more samples than others. To address this, class weights

Table 2. *Performance results for all scenarios*

Algorithms		Performance Evaluation Criteria		
		Accuracy	F_1 Weighted	Roc-Auc Ovo Weighted
PCA	MLP	0.49	0.36	0.55
	LR	0.33	0.28	0.60
	KNN	0.45	0.42	0.53
	DT	0.45	0.44	0.56
	RF	0.48	0.44	0.60
	Adaboost	0.41	0.36	0.57
	GBM	0.45	0.43	0.56
	XGBoost	0.45	0.43	0.54
	LightGBM	0.41	0.40	0.57
LDA	MLP	0.52	0.41	0.58
	LR	0.34	0.27	0.61
	KNN	0.44	0.40	0.56
	DT	0.34	0.33	0.48
	RF	0.44	0.40	0.56
	Adaboost	0.47	0.40	0.58
	GBM	0.43	0.41	0.55
	XGBoost	0.41	0.40	0.53
	LightGBM	0.37	0.36	0.58
Before Hyperparameter Optimization	MLP	0.51	0.35	0.52
	LR	0.29	0.23	0.56
	KNN	0.45	0.41	0.54
	DT	0.30	0.29	0.46
	RF	0.45	0.42	0.56
	Adaboost	0.42	0.38	0.55
	GBM	0.46	0.45	0.56
	XGBoost	0.44	0.43	0.58
	LightGBM	0.38	0.37	0.57
After Hyperparameter Optimization	KNN	0.52	0.45	0.59
	DT	0.49	0.47	0.61
	RF	0.47	0.40	0.65
	Adaboost	0.46	0.44	0.58
	GBM	0.55	0.53	0.68
	XGBoost	0.47	0.51	0.62
	LightGBM	0.52	0.47	0.66
	Hybrid(DT+GBM+XGB)	**1.00**	**1.00**	**1.00**

were incorporated into the Decision Tree's Gini Index, emphasizing minority classes and improving model robustness [30]. Lactation milk yield was classified into low, medium, and high, with higher weights assigned to underrepresented groups.

Following the training and hyperparameter tuning procedures described in previous sections, the models were evaluated on three different datasets: (a) the original feature set, (b) features reduced via Principal Component Analysis (PCA), and (c) features reduced via Linear Discriminant Analysis (LDA). Performance results for these scenarios, obtained via 10-fold cross-validation, are presented in *Table 2*.

The results obtained through PCA and LDA indicate that most models experienced either a slight decrease in performance or yielded comparable outcomes. This suggests that reducing the feature space to two components did not significantly enhance model learning. This finding is expected, as the original dataset contained only five features; thus, reducing it to two components may have resulted in some information loss. Nevertheless, the performance differences were generally minör for instance, the Random Forest (RF) model achieved a weighted F1 score of 0.42 on the original dataset and 0.40 with LDA. These analyses provided valuable insights into the impact of dimensionality reduction on model performance. At this stage, the models achieving the highest F1 scores were GBM, XGBoost, and RF. The classification accuracies of the hybrid model developed using the bootstrap sampling method were compared with those of DT, GBM, and XGBoost. For this analysis, 50 bootstrap samples were generated, each containing 10 observations. Under these bootstrap sampling conditions, the accuracy scores of DT, GBM, XGBoost, and the hybrid model were evaluated. As a result, the average accuracy and weighted F1 score were obtained as 1.00.

Fig. 3 illustrates the distribution of weighted F1 scores for individual models across bootstrap samples. While the Decision Tree (DT) model shows considerable fluctuation, the other models and the hybrid model consistently achieved a perfect score, exhibiting the highest and most stable performance across all bootstrap subsets. The same experiment was repeated to verify accuracy, and identical results were obtained.

The proposed hybrid model's weighted ROC-AUC scores were evaluated across all possible pairwise class combinations using a One-vs-One (OvO) approach to provide a more detailed assessment of its performance.

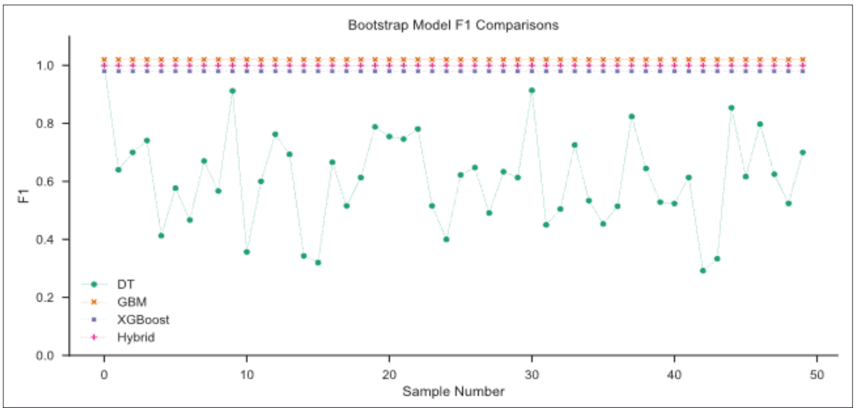

Fig 3. F1 comparisons for bootstrap sampling (10 observations * 50 samples)

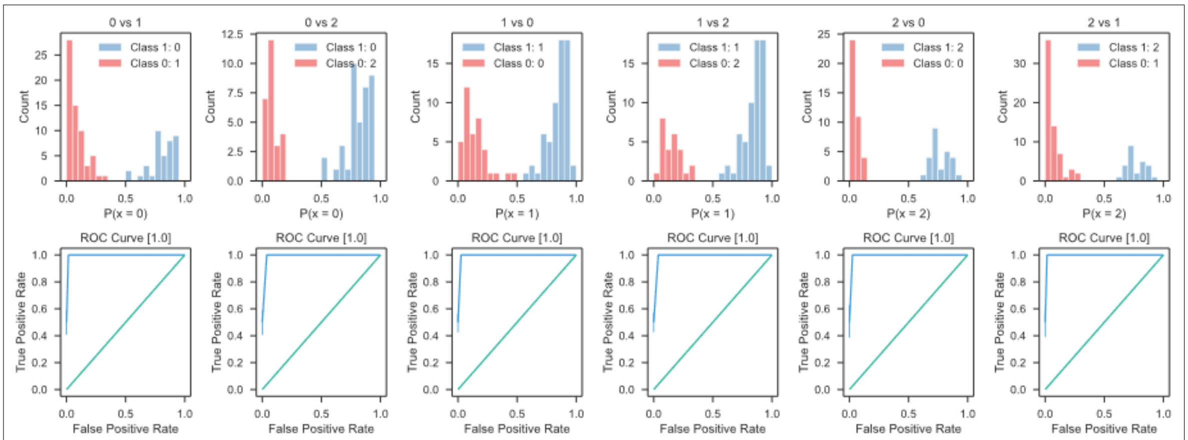

Fig 4. ROC-AUC (One vs One) for the Hybrid Model

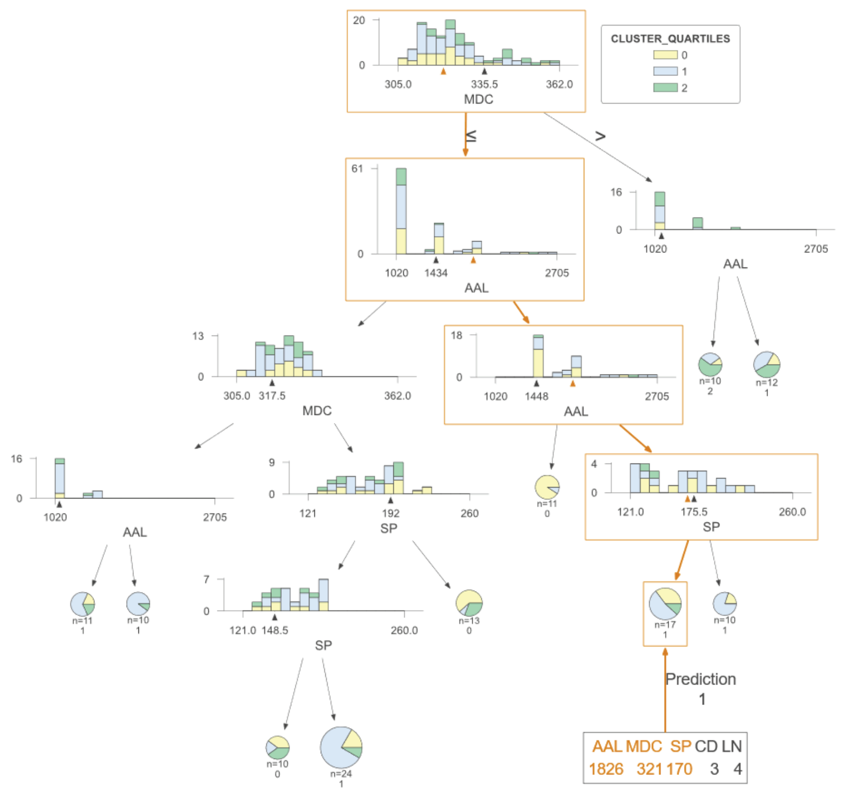

Fig 5. Visualization of the classification result for the 41st observation using the dtreeviz library and the detailed presentation of the decision path

These analyses offer a more comprehensive understanding of the model's ability to discriminate between classes, highlighting its effectiveness in multiclass classification scenarios. The AUC score for all possible pairs was 1.00. Visualizations of these findings are presented in *Fig. 4*.

Variable importance scores were calculated to identify the most influential variables in the decision-making process of the Decision Tree (DT), Gradient Boosting Machine (GBM), and XGBoost models, which yielded the best scores after hyperparameter optimization. Service Period (SP) was identified as the most important variable for the DT model, Animal's Age in Lactation (ALL) for the GBM model, and Milking Days Count (MDC) for the XGBoost model. These findings are intuitively significant, as longer milking durations and appropriate insemination intervals are generally associated with higher total milk production.

In addition, to analyze how the model makes individual predictions, a detailed examination of the decision path was conducted. In this context, the classification path followed by the Decision Tree model for cow number 41, which was selected to evaluate classification accuracy and overall model performance, is presented in *Fig. 5*. This approach serves as a crucial example for understanding the model's decision-making process more transparently. Such individual case studies reveal not only the statistical performance of the model but also its interpretability for practical field applications [25].

The hybrid ensemble model, combining GBM, XGBoost, and DT via majority voting, outperformed all individual models, raising the best single-model F1-score (GBM: 0.53) to 1.00. This reflects the complementary strengths of tree-based and boosting methods, producing a robust classification framework. Cross-validation and bootstrap resampling helped mitigate overfitting, though the small, imbalanced dataset remains a limitation.

The study aimed to classify Holstein cows' milk yield (low, medium, high) using five features from 128 samples. Preprocessing included scaling, outlier removal, and correlation analysis. PCA and LDA were tested but discarded due to no performance gain. Nine algorithms were evaluated with weighted F1-score as the main metric. Boosting-based methods performed best after hyperparameter tuning, but the hybrid model achieved perfect separation, offering a reliable decision-support tool under similar constraints.

Discussion

This study demonstrated that the lactation milk yield levels (low, medium, high) of Holstein cows can be classified with high accuracy using supervised machine learning algorithms. In particular, the hybrid model integrating the strengths of Decision Tree (DT), Gradient

Boosting Machine (GBM), and XGBoost achieved outstanding classification performance, with a weighted F1-score of 1.00. This result underscores the effectiveness of ensemble methods, even in the presence of challenges such as limited data and class imbalance. Previous studies have reported that boosting-based algorithms are capable of building robust and reliable models by reducing the risk of overfitting, especially in small and complex datasets [41-43]. In this study, the inclusion of biologically meaningful features such as age, number of lactations, insemination interval, and number of milking days played a crucial role in model performance. Proper integration of relevant biological and management-related factors directly enhanced classification accuracy. These findings emphasize that in cases of limited data, careful feature selection and preparation significantly improve model performance. Despite the pronounced class imbalance within the dataset, the model delivered high accuracy. In particular, the correct classification of low-yield cows was supported by highly sensitive performance metrics such as the weighted F1-score and ROC-AUC. The elimination of outliers, as well as preprocessing steps like data standardization and hyperparameter optimization, contributed significantly to reducing the negative impact of class imbalance. This finding confirms that appropriate data handling and modeling techniques can effectively address the challenges posed by imbalanced datasets [44-46].

This study makes a significant contribution to the development of decision support systems at the individual animal level. While previous research has primarily focused on farm-level management strategies using unsupervised learning techniques [13], the present study employs supervised algorithms to classify the lactation yield level of each cow individually. This enables the implementation of personalized, producer-specific management decisions. Furthermore, the use of a hybrid model that combines the complementary strengths of different algorithms has overcome the limitations of single-model approaches, resulting in more stable, reliable, and generalizable outcomes [47,48].

The study has several important limitations. First, the dataset used was exclusively composed of Holstein cattle from a single farm located in one geographic region. This restricts the generalizability of the model to other breeds (e.g., Jersey, Brown Swiss, Guernsey, Milking Shorthorn) or different environmental and management conditions. As highlighted in the literature, machine learning models often require retraining and validation to ensure their applicability across diverse environmental and operational contexts [49,50]. Secondly, the developed model was designed solely for classification purposes and does not provide quantitative predictions of milk yield. Employing regression-based models could offer more functional insights for production planning by enabling direct estimation of milk output [51,52]. Lastly, the model focuses exclusively on milk yield and does not incorporate other key production parameters such as fertility, feed intake, and health status. Integrating these factors would enable the development of a more comprehensive and holistic decision support system [53,54].

The integration of data-driven decision support systems into animal production processes not only enhances production efficiency but also contributes to the development of sustainable management approaches that prioritize animal welfare [55,56]. Advanced studies in this direction will facilitate the real-time application of data-based models, enabling production processes to be managed in a more traceable, optimized, and ethically grounded manner. Consequently, this will lead to scientific and practical solutions aimed at improving both economic performance and ensuring animal health and welfare.

This study demonstrated that machine learning methods can effectively classify lactation milk yield levels in Holstein cows with high accuracy, even under limited and imbalanced data conditions. The superior performance of the hybrid model highlights the critical role of carefully selected biologically relevant variables and comprehensive data preprocessing strategies in model success. The findings support the development of individual animal-based, data-driven decision support systems and emphasize the potential of artificial intelligence applications to enhance both productivity and animal welfare in livestock farming. However, the model's reliance on data from a single breed and a geographically limited region presents a constraint regarding its generalizability. Future research should focus on validating the model across different breeds and environmental conditions, expanding it with regression-based approaches for quantitative yield prediction, and integrating additional key production parameters such as health, fertility, and feed intake.

Such advancements would facilitate the integration of data-driven AI applications into real-time, sustainable, and ethical livestock management systems ultimately contributing to improved economic performance and enhanced animal welfare.

Declarations

Availability of Data and Materials: Data and materials for this research are available upon request.

Acknowledgments: The corresponding author would like to express his heartfelt gratitude to his graduate student for the unwavering support during this research. The authors would also like to thank the reviewers for their careful, constructive, and insightful comments regarding this work.

Funding: This research did not receive any specific grant from

funding.

Ethical Statement: This study does not require ethical permission.

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

Declaration of Generative Artificial Intelligence (AI): The authors declare that the article and/or tables and figures were not written/created by AI and AI-assisted technologies.

Author Contributions: DT and ST contributed to the design of this study. DT and ST participated in the sample collection, data analysis. DT and ST wrote the original draft. All authors contributed to data collection and discussion.

References

1. Cozzi G, Brscic M, Gottardo F: Animal welfare as a pillar of a sustainable farm animal production. *Acta Agric Slov,* 2, 23-31, 2008. DOI: 10.14720/aas-s.2008.2.19200

2. Bell MJ, Wall E, Russell G, Roberts DJ, Simm G: Risk factors for culling in Holstein-Friesian dairy cows. *Vet Rec,* 167 (7): 238-240, 2010. DOI: 10.1136/vr.c4267

3. Bahoo S, Cucculelli M, Qamar D: Artificial intelligence and corporate innovation: A review and research agenda. *Technol Forecast Soc Change,* 188:122264, 2023. DOI: 10.1016/j.techfore.2022.122264

4. Rutten CJ, Velthuis AGJ, Steeneveld W, Hogeveen H: Invited review: Sensors to support health management on dairy farms. *J Dairy Sci,* 96 (4): 1928-1952, 2013. DOI: 10.3168/jds.2012-6107

5. Rorie RW, Bilby TR, Lester TD: Application of electronic estrus detection technologies to reproductive management of cattle. *Theriogenology,* 57 (1): 137-148, 2002. DOI: 10.1016/S0093-691X(01)00663-X

6. Firk R, Stamer E, Junge W, Krieter J: Automation of oestrus detection in dairy cows: A review. *Livest Prod Sci,* 75 (3): 219-232, 2002. DOI: 10.1016/S0301-6226(01)00323-2

7. Fatima M, Pasha M: Survey of machine learning algorithms for disease diagnostic. *J Intell Learn Syst Appl,* 9 (1): 1-16, 2017. DOI: 10.4236/jilsa.2017.91001

8. 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

9. Koklu M, Ozkan IA: Multiclass classification of dry beans using computer vision and machine learning techniques. *Comput Electron Agric,* 174:105507, 2020. DOI: 10.1016/j.compag.2020.105507

10. Cihan P, Gökçe E, Kalıpsız 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

11. Ozger ZB, Cihan P, Gokce E: A Systematic review of IoT technology and applications in animals. *Kafkas Univ Vet Fak Derg,* 30 (4): 411-431, 2024. DOI: 10.9775/kvfd.2024.31866

12. Cihan P, Saygılı A, Özmen NE, Akyüzlü M: Identification and recognition of animals from biometric markers using computer vision approaches: A review. *Kafkas Univ Vet Fak Derg,* 29 (6): 581-593, 2023. DOI: 10.9775/kvfd.2023.30265

13. Brotzman RL, Cook NB, Nordlund K, Bennett TB, Rivas AG, Döpfer D: Cluster analysis of dairy herd ımprovement data to discover trends in performance characteristics in large upper midwest dairy herds. *J Dairy Sci,* 98 (5): 3059-3070, 2015. DOI: 10.3168/jds.2014-8369

14. Van Klompenburg T, Kassahun A, Catal C: Crop yield prediction using machine learning: A systematic literature review. *Comput Electron Agric,* 177 (10):105709, 2020. DOI: 10.1016/j.compag.2020.105709

15. Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D: Machine learning in agriculture: A review. *Sensors,* 18 (8):2674, 2018. DOI: 10.3390/s18082674

16. 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

17. Curti PDF, Selli A, Pinto DL, Merlos-Ruiz A, Balieiro JCDC, Ventura RV: Applications of livestock monitoring devices and machine learning algorithms in animal production and reproduction: An overview. *Anim Reprod,* 20 (2):e20230077, 2023. DOI: 10.1590/1984-3143-AR2023-0077

18. 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

19. Caraviello DZ, Weigel KA, Craven M, Gianola D, Cook NB, Nordlund KV, Fricke PM, Wiltbank MC: Analysis of reproductive performance of lactating cows on large dairy farms using machine learning algorithms. *J Dairy Sci,* 89 (12): 4703-4722, 2006. DOI: 10.3168/jds.S0022-0302(06)72521-8

20. Shine P, Murphy MD, Upton J, Scully T: Machine-learning algorithms for predicting on-farm direct water and electricity consumption on pasture based dairy farms. *Comput Electron Agric,* 150, 74-87, 2018. DOI: 10.1016/j.compag.2018.03.023

21. Breunig MM, Kriegel HP, Ng RT, Sander J: LOF: Identifying density-based local outliers. *In Proc Int Conf Manag Data,* 93-104, 2000. DOI: 10.1145/335191.335388

22. Nilsson NJ: Introduction to machine learning. An early draft of a proposed textbook, Robotic Library. Department of Computer Science, Stanford University, 1998.

23. Liu Y, Chen B, Qiao J: Development of a machine vision algorithm for recognition of peach fruit in natural scene. *Trans ASABE,* 54 (2): 695-702, 2011. DOI: 10.13031/2013.36472

24. Takma Ç, Atıl H, Aksakal V: Çoklu doğrusal regresyon ve yapay sinir ağı modellerinin laktasyon sut verimlerine uyum yeteneklerinin karşılaştırılması. *Kafkas Univ Vet Fak Derg,* 18 (6): 941-944, 2012. DOI: 10.9775/kvfd.2012.6764

25. Breiman L: Bagging predictors. *Mach Learn,* 24 (2): 123-140, 1996. DOI: 10.1007/BF00058655

26. Sarker IH: Machine learning: Algorithms, real-world applications and research directions. *SN Comput Sci,* 2 (3): 160, 2021. DOI: 10.1007/s42979-021-00592-x

27. Torlay L, Perrone-Bertolotti M, Thomas E, Baciu M: Machine learninge XGBoost analysis of language networks to classify patients with epilepsy. *Brain Inform,* 4 (3): 159-169, 2017. DOI: 10.1007/s40708-017-0065-7

28. Şanlı T, Sıcakyüz Ç, Yüregir OH: 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. Hempstalk K, McParland S, Berry DP: Machine learning algorithms for the prediction of conception success to a given insemination in lactating dairy cows. *J Dairy Sci,* 98 (8): 5262-5273, 2015. DOI: 10.3168/jds.2014-8984

30. Gonzalez-Recio O, Weigel KA, Gianola D, Naya H, Rosa GJ: L2-boosting algorithm applied to high-dimensional problems in genomic selection. *Genet Res,* 92 (3): 227-237, 2010. DOI: 10.1017/S0016672310000261

31. Fawagreh K, Gaber MM, Elyan E: Random forests: from early developments to recent advancements. *Syst Sci Contr Eng,* 2 (1): 602-609, 2014. DOI: 10.1080/21642583.2014.956265

32. Zhu M, Xia J, Jin X, Yan M, Cai G, Yan J, Ning G: Class weights random forest algorithm for processing class imbalanced medical data. *IEEE Access,* 6, 4641-4652, 2018. DOI: 10.1109/ACCESS.2018.2789428

33. Jia BB, Zhang ML: Multi-dimensional classification via KNN feature augmentation: Pattern recognition. *Assoc Adv Artif Intell,* 106:107423, 2020. DOI: 10.1016/j.patcog.2020.107423

34. James G, Witten D, Hastie T, Tibshirani R: An introduction to statistical learning*:* 103 Springer Texts in Statistics New York: Springer, 2013. DOI: 10.1007/978-1-4614-7138-7

35. Sharma R, Kamble SS, Gunasekaran A, Kumar V, Kumar A: A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. *Comput Oper Res,* 119:104926, 2020. DOI: 10.1016/j.cor.2020.104926

36. Lipton ZC, Elkan C, Narayanaswamy B: Optimal thresholding of classifiers to maximize F1 measure. *Mach Learn Knowl Discov Databases,* 8725, 225-239, 2014. DOI: 10.1007/978-3-662-44851-9_15

37. Shine P, Scully T, Upton J, Murphy MD: Multiple linear regression modelling of on-farm direct water and electricity consumption on pasture based dairy farms. *Comput Electron Agric,* 148, 337-346, 2018. DOI: 10.1016/j.compag.2018.02.020

38. Jaskowiak PA, Costa IG, Campello RJ: The area under the ROC curve as a measure of clustering quality. *Data Mining Knowl Discov,* 36 (3): 1219-1245, 2022. DOI: 10.1007/s10618-022-00829-0

39. Shine P, Scully T, Upton J, Shalloo L, Murphy MD: Electricity and direct water consumption on Irish pasture based dairy farms: A statistical analysis. *Appl Energy,* 210, 529-537, 2018. DOI: 10.1016/j.apenergy.2017.07.029

40. Sahinler S, Topuz D: Bootstrap and jackknife resampling algorithms for estimation of regression parameters. *JAQM,* 2 (2): 188-199. 2007.

41. Chen T, Guestrin C: Xgboost: A scalable tree boosting system. **In,** *Proceedings of the 22nd acm Sigkdd İnternational Conference on Knowledge Discovery and Data Mining,* San Francisco, 13-17 August, 785-794, 2016. DOI: 10.1145/2939672.2939785

42. Probst P, Wright MN, Boulesteix AL: Hyperparameters and tuning strategies for random forest. *WIREs Data Mining Knowl Discov,* 9 (3):e1301, 2019. DOI: 10.1002/widm.1301

43. Natekin A, Knoll A: Gradient boosting machines, a tutorial. *Front Neurorobot,* 7 (21): 1-21. 2013. DOI: 10.3389/fnbot.2013.00021

44. Jafarigol E, Trafalis TA: A review of machine learning techniques in imbalanced data and future trends. ArXiv Preprint, 2023. DOI: 10.48550/arXiv.2310.07917

45. Haibo H, Garcia EA: Learning from imbalanced data. *IEEE Trans Knowl Data Eng,* 21 (9): 1263-1284, 2009. DOI: 10.1109/TKDE.2008.239

46. Buda M, Maki A, Mazurowski MA: A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw,* 106, 249-259, 2018. DOI: 10.1016/j.neunet.2018.07.011

47. Zhao ZQ, Zheng P, Xu ST, Wu X: Object detection with deep learning: A review. *IEEE Trans Neural Netw Learn Syst,* 30 (11): 3212-3232, 2019. DOI: 10.1109/TNNLS.2018.2876865

48. Dietterich TG: Ensemble methods in machine learning. In: Multiple Classifier Systems, Lecture Notes in Computer Science, 1857:1-15, Springer, Berlin, Heidelberg, 2000. DOI: 10.1007/3-540-45014-9-1

49. Parivendan SC, Sailunaz K, Neethirajan S: Socializing AI: Integrating social network analysis and deep learning for precision dairy cow monitoring. A critical review. *Animals,* 15 (13): 1835, 2025. DOI: 10.3390/ani15131835

50. Rios J: El acuerdo de paz entre el gobierno colombiano y las FARC: o cuando una paz imperfecta es mejor que una guerra perfecta. Araucaria Rev Iberoam Filos Polit Humanid, 19(38): 593-618, 2017. DOI: 10.12795/araucaria.2017.i38.28

51. Gonzalez-Recio O, Rosa GJ, Gianola D: Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. *Livest Sci,* 166, 217-231, 2014. DOI: 10.1016/j.livsci.2014.05.036

52. de Los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MP: Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics,* 193 (2): 327-345, 2013. DOI: 10.1534/genetics.112.143313

53. Pryce JE, Veerkamp RF: The incorporation of fertility indices in genetic improvement programs. *BSAS Occas Publ Fert,* 26 (1): 237-249, 2001. DOI: 10.1017/S0263967X00033711

54. Berry HL, Hogan A, Owen J, Rickwood D, Fragar L: Climate change and farmers mental health: Risks and responses. *Asia Pac J Public Health,* 23 (2): 119-132, 2011. DOI: 10.1177/1010539510392556

55. Wolfert S, Ge L, Verdouw C, Bogaardt MJ: Big data in smart farming - A review. *Agric Syst,* 153, 69-80, 2017. DOI: 10.1016/j.agsy.2017.01.023

56. Eastwood C, Klerkx L, Ayre M, Dela Rue B: Managing socio-ethical challenges in the development of smart farming: From a fragmented to a comprehensive approach for responsible research and innovation. *J Agric Environ Ethics,* 32 (5): 741-768, 2019. DOI: 10.1007/s10806-017-9704-5