Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 6, 2026

A MACHINE LEARNING FRAMEWORK FOR EARLY DETECTION AND DIAGNOSIS OF CANINE DIABETES

Fareed Ahmad[1,2,*], Muhammad Usman Ghani Khan[1], Irfan Irshad[3], Muhammad Yasin Tipu, [3]Muhammad Munwar Iqbal[5]

[1]Department of Computer Science, University of Engineering and Technology, Lahore, Pakistan
[2]Quality Operations Laboratory, Institute of Microbiology, University of Veterinary and Animal Sciences, Lahore, Pakistan
[3]Department of Continuing Education, University of Engineering and Technology, Lahore, Pakistan
[4]Department of Pathology, University of Veterinary and Animal Sciences, Lahore, Pakistan
[5]Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan
fareed.ahmad@uvas.edu.pk, usman.ghani@uet.edu.pk, irfanirshad@uvas.edu.pk, yasin.tipu@uvas.edu.pk, munwariq@gmail.com

DOI:- https://doi.org/10.5281/zenodo.20791800

Keywords
Canines, Diabetes, Machine learning Algorithms, Classification, Feature ranking

Article History
Received: 12 May, 2026
Accepted: 16 June, 2026
Published: 17 June, 2026

Corresponding Author: *
Fareed Ahmad

Abstract
Disease diagnosis in animals is often more challenging than in humans due to their inability to communicate their symptoms directly, and because many diseases exhibit similar clinical signs. Predicting the risk of diabetes in dogs is difficult because veterinary datasets are often noisy, imbalance, and contains heterogeneous clinical measurements. Machine learning based decision support systems offer an effective approach for analyzing such complex data to improve disease diagnostic accuracy. These systems can assist veterinary care providers in maintaining round-the-clock remote surveillance and enable veterinarians to have instant access to relevant patient information. This study presents a Canine Diabetes Diagnosis and Recommendation (CDDR) framework for predicting the severity of diabetes in canines using machine learning classifiers. Information Gain, a feature selection method, is used to identify the most relevant clinical and laboratory features, thereby reducing data dimensionality and improving model performance. Several machine learning algorithms, including Random Forest, LibSVM, Decision Stump, and REP Tree, were evaluated using 10-fold cross-validation. Among the evaluated classifiers within the CDDR framework, Random Forest achieves the highest accuracy of 93.0%, precision of 0.92, recall of 0.92, and the lowest mean absolute error of 0.07. Overall, our findings indicate that integrating feature selection, machine learning techniques, and decision support systems can significantly improve the accuracy and reliability of canine diabetes prediction. The proposed CDDR framework can assist veterinarians in early disease detection, clinical decision-making, and continuous remote monitoring of canine patients, especially in resource-constrained veterinary settings.

Introduction

Diabetes mellitus is one of the chronic diseases in humans. The worldwide occurrence of diabetes across different age groups was estimated to be 2.8% in 2000 and 4.4% in 2030. Furthermore, in 2014, the global prevalence of diabetes was 9% among adults over 18 years of age [1,2]. However, the disease would surge to 53% between 2014-35, increasing from 387 to 592 million people [3,4]. The situation is much more alarming in US where almost 29.1 million people are suffering from diabetes, which is estimated to be 9.3% of the population [5,6]. In case of pre-diabetes, the situation is even more concerning, as 86 million adults are affected and 90% are ignorant about it [7,8]. Almost 69,071 people in US lost their lives due to diabetes in 2010. This makes it the 7th leading cause of death in the United State [9,10]. Annually, it was estimated that 5 million people die directly due to diabetes over the globe [11,12]. The diabetes has increased from 7% in 1990 to 14% in 2022. The reports further indicate that, in canines, the disease has increased by approximately 12.8% annually [13]. The occurrence of diabetes in cats and dogs ranges from approx. 0.4–1.2% and during the last 3 decades number of dogs diagnosed with diabetes has increased three folds [14,15]. It is an endocrine disorder in dogs and certain breeds are at greater risk of developing the disease than others. Epidemiologic studies have also revealed that most of the dogs affected with diabetes are greater than five years of age [16,17]. It has also been reported in several studies that female dogs have an increasing risk of the disease [18,19]. In UK the prevalence of the disease is estimated as 0.32–0.36%. Dogs show almost similar clinical signs, as in humans, including weight loss, polyuria and polydipsia, related with glycosuria and hyperglycemia [20]. Some studies suggest that in America, during 2006 to 2015 the number of diabetic canines increased from 13.1 to 23.6 per 10,000, thereby reflecting that the disease is increasing in dogs as in humans [21].

The recent developments in machine learning and artificial intelligence have opened a new era for researchers in disease diagnosis in human health. These techniques can be used as powerful tools in healthcare for analyzing complex and large datasets in the field of disease diagnosis [22]. Machine learning and deep learning algorithms have the ability to learn patterns from large amounts of medical and veterinary data, made available through the advent of large and efficient storage devices on local and remote data servers. These algorithms can help medical and veterinary professionals in better identification of risk factors and disease prediction, resulting in an efficient and reliable decision-making system that could lead to better healthcare services for the community [23,24].

Machine and deep learning techniques have shown exceptional results in a wide range of healthcare issues, including cancer diagnosis [25], medical image analysis [26], prediction of heart disease [27], and infectious disease detection. These models are applied for risk assessment and diagnosis of diabetes in humans, which can assist medical practitioners in the detection of the disease, thereby reducing disease complications [28,29].

Decision Support Systems (DSS) in the field of medicine are computer-based tools that can assist healthcare professional in clinical decision making by combining medical data with medical knowledge. Modern DDS use machine learning to help with diagnosis, assess risks, suggest treatments, and provide recommendations based on each patient's clinical lesions [30, 31]. These systems can also help specialist in making accurate and consistent clinical decision making for patients by screening and predicting risks, planning treatment and monitoring outcomes. The integration of machine learning with decision support systems can provide a foundation for intelligent healthcare systems.

Although machine learning is currently being applied for disease diagnosis in medical healthcare, the field of veterinary medicine is largely ignored. The development of intelligent clinical decision support system for canine diabetes prediction remains limited due to resource constraints in veterinary settings. Therefore, we aim to present a machine learning-based Canine Diabetes Diagnosis and Recommendation (CDDR) framework for canine diabetes prediction. The framework support veterinary decision making by integrating data preprocessing, feature selection, classification, risk assessment, and recommendation under a single umbrella. Different classification models were applied in the framework to identify the most effective model for prediction of the disease. Our proposed framework intends to assist veterinarians in early diagnosis and management of diabetes in canines.

The remainder of this paper is organized as follows. Section 2 presents the related work, Section 3 describes the methodology, Section 4 discusses the experimental results and analysis, and Section 5 concludes the study.

1 Related Work

Machine learning models and deep learning models have been widely used in healthcare for the prediction and diagnosis of diabetes. Various studies have applied these algorithms to identify diabetic patients using clinical and demographic data. Elaheh et al. [28] applied various machine learning models for diabetes diagnosis and evaluated their results using performance metrics such as accuracy, precision, recall, and F1-score metrics. The study applied Random Forest and achieve an accuracy of 82.26%. Karunakaran, D. et al. [29] proposes an intelligent remote monitoring system for surveillance of patients with diabetes using smartphone and portable sensors. An attention-based recurrent unit was applied to detect critical changes in patient parameters. Various evaluation metrics, including F-score, accuracy, error rate etc. were applied to access the performance of different machine models against traditional approaches.

Machine learning has achieved considerable success in human healthcare, but its application in veterinary sector remains limited. Recent studies have applied machine learning models for disease diagnosis and prediction in animals, as well as livestock and poultry management [32,33,34]. These studies reflect that such intelligent systems can help veterinarians in disease diagnosis and decision making. Soumen et al. [35] investigates the impact of various animal disease on livestock production, which highlights the importance data-driven solutions used in machine learning models in the field of veterinary science and agriculture for disease prediction and analysis. The study explorers both supervised and unsupervised models on veterinary dataset to improve accuracy and decision making in animal healthcare. Swain et al. [36] utilize various machine learning techniques to predict various diseases in cattle. The research depicts that random forest outperformed other machine learning models (NBM, IBk, PART, and SVM) by achieving an accuracy of 88%. Zhou et al. [37] used high-dimensional data from automated milking systems for disease prediction in dairy cows. The study initially applied eight machine learning models to various dairy parameters such as season, days in milking, parity, age at disorder onset, and milk yield. The model achieved an accuracy of 81.58% and a precision of 92.86% using Rpart, which reflects that these models can help in establishing a decision support and monitoring system for the detection of health disorders in animal flocks.

Decision support systems are being used increasingly in healthcare for diagnosis, treatment, and disease management. These systems apply machine learning models for recommendation and decision support using domain knowledge and patient data. Numerous decision support systems have been developed for the diagnosis of diabetes in patients. There are some examples of these systems for disease diagnosis in cats and dogs [30,31]. However, to best of our knowledge, there is no decision support framework that applies machine learning for canine diabetes prediction. Furthermore, there is limited research comparing the performance of different machine learning classifiers for this task. This study addresses this gap by developing the Canine Diabetes Diagnosis and Recommendation (CDDR) framework and evaluating multiple established machine learning classifiers to identify the most suitable prediction model for canine diabetes.

2 Materials and Methods

2.1 Overview of the CDDR Framework

A proposed Canine Disease Diagnosis and Recommendation (CDDR) framework that applies machine learning models for recommendation and decision support is presented in Figure 1. This decision support framework can assist veterinarians in the diagnosis of canine diabetes. The framework consists of three main steps: In the first layer demographic, clinical, and laboratory data is collected and preprocessed by handling missing values, encoding categorical variable, and data transformation. The relevant attributes are then selected by applying the information gain feature selection technique. In the second layer the selected features are used to train machine learning classification models. A 10-folds cross validation is applied and the best classifier is selected to generate diabetes diagnosis and risk predictions. In the final layer, recommendations are generated for predicted results. These recommendations and predictions are then vetted by veterinary practitioners, and their feedback is incorporated through a validation loop to improve the quality and reliability of future recommendations.

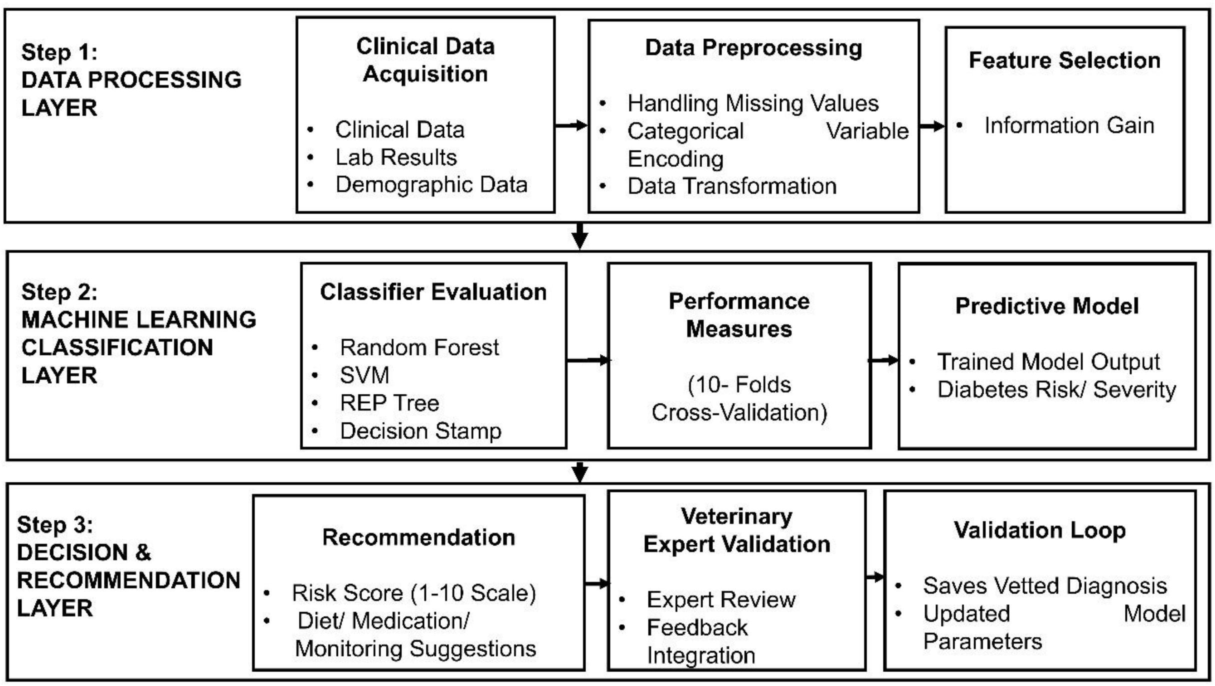

Figure 1. Overview of the Canine Disease Diagnosis Recommender (CDDR) framework.

2.2 Dataset Collection

The first objective was to gather data that is necessary for analysis of diabetes in canines. The relevant data was gathered from pet centers, veterinary hospitals. The gathered data record consists of number of attributes, values and diagnosis by veterinarians in each case. The relevant veterinary patient information was entered through a structured data collection form used to construct the canine diabetes dataset, as shown in Figure 2.

Relevant veterinary textbooks, clinical pathology resources, and research articles were reviewed to identify the most important risk factors and clinical indicators associate with diabetes in canines. In addition, veterinary practitioners were also consulted for the validation of the selected attributes. The final dataset comprises 181 records pertaining to canines with different signs and symptoms. Each record in the dataset consists of clinical symptoms, laboratory results, demographic data and final diagnosis made

Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 6, 2026

by veterinary practitioner based on these factors. The dataset is used to train machine learning models in order to predict diabetes in canines.

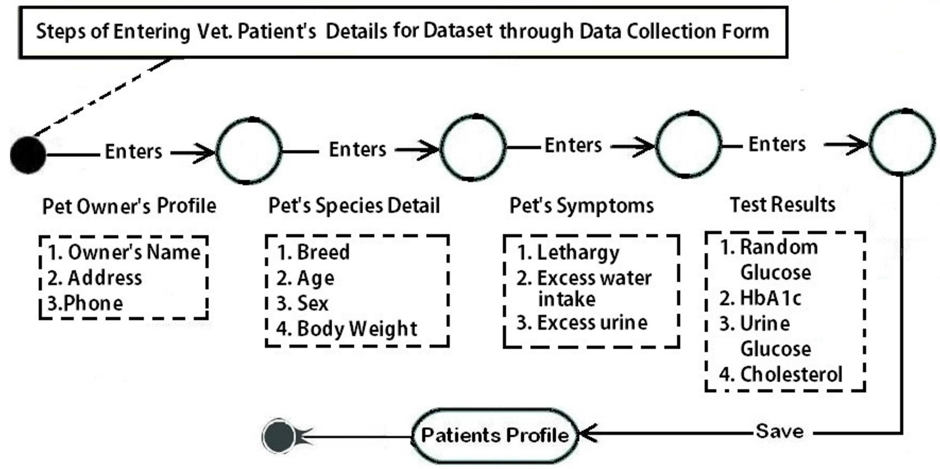

Figure 2. Structured data collection form used for recording patient cases for canine diabetes dataset.

2.3 Clinical Attributes

Different types of attributes were selected such as demographic data, physiological data, laboratory results and sign and symptoms associated with diabetes in canines. These risk factors include age, gender, breed susceptibility, and body weight status. Age was selected because diabetes is more common in older dogs (5–15 years) than in younger dogs (< 5 years), while breed and gender were included because some breeds are more susceptible to the disease, and female dogs have a higher risk of developing the disease than males [38-40]. Random glucose, HbA1C, and Serum cholesterol are also strong indicators of diabetes and were therefore included as laboratory parameters [41,42]. Excessive urination (polyuria), excessive water intake (polydipsia), presence of glucose in urine (glycosuria), and body weight status are clinical signs consistently associated with canine diabetes. The details of these critical indicators is presented in the Table 1.

2.4 Data Preprocessing

Before training the machine learning classifiers within the CDDR framework, the collected data must be preprocessed. Preprocessing improves the quality of the data and classification performance. The initial step in preprocessing is to handle incomplete or inconsistent values. These values were corrected through verification of veterinary records. Next step is to identify the most important features using Information Gain. This feature selection method ranks features according to their contribution towards predicting the class variable. After selecting appropriate features that data needs to be transformed into a suitable format for machine learning models. This processing step reduces data redundancy and improves accuracy. A sample of the dataset for

Figure 3. Sample records from the 181-case canine diabetes dataset.

VetPatientNo	Age	Sex	Breed	Weight	RBS	HbA1c	Lethargy	Excess_Urine	Excess_Water_Intake	Urine_Glucose	Cholesterol	Diagnosed_Class
78801	1	male	Boxers	20	75	2.4	FALSE	FALSE	FALSE	FALSE	156	1
78802	2	male	German Shephard	29	95	2.6	FALSE	FALSE	FALSE	FALSE	150	3
78803	11	female	Boxers	22	230	8	TRUE	TRUE	TRUE	TRUE	416	9
78804	4.5	male	German Shephard	30	146	3.7	FALSE	FALSE	FALSE	FALSE	260	6
78805	1.5	male	Labrador	26	223	5.2	TRUE	TRUE	TRUE	TRUE	280	8
78806	3	female	Boxers	24	136	3.5	FALSE	FALSE	FALSE	FALSE	240	5
78807	1.25	male	Minature Poodles	4	112	3.1	FALSE	FALSE	FALSE	FALSE	198	4
78808	12	female	Doberman	32	410	9	TRUE	TRUE	TRUE	TRUE	367	10
78809	2	male	German Shephard	25	114	3	FALSE	FALSE	FALSE	FALSE	174	4
78810	5.5	female	Labrador	29	300	7.8	TRUE	TRUE	TRUE	TRUE	290	9
78811	3	male	Minature Poodles	6.5	140	3.7	FALSE	FALSE	FALSE	FALSE	210	6
78812	2	female	Keeshonds	27	95	2.8	FALSE	FALSE	FALSE	FALSE	246	2
78813	2	male	Doberman	33	85	2.6	FALSE	FALSE	FALSE	FALSE	246	1
78814	10	male	Bull dog	29	160	4.8	TRUE	TRUE	TRUE	FALSE	250	7
78815	5	female	German Shephard	30	126	3.2	FALSE	FALSE	FALSE	FALSE	165	4

diabetes in canines is shown in the Figure 3.

Table 1. Parameters, Values, and Prevalence for Canine Diabetes Analysis

Parameters	Value	Prevalence
Age	Age < 5	Less Prevalent
	Age ≥ 5 to ≤ 15	More Prevalent
	Age > 15	(Not specified)
Gender	Male	Less Prevalent
	Female	More Prevalent
Breed	Collies	Less Likely
	Boxers	Less Likely
	German Shepherds	Less Likely
	Cocker Spaniels	Less Likely
	Keeshonds	More Likely
	Alaskan Malamutes	More Likely
	Finnish Spitz	More Likely
	Miniature Schnauzers	More Likely
	Miniature Poodles	More Likely
	English Springer Spaniels	More Likely
Weight	Underweight	More Prevalent
	Normal	Less Prevalent
Random Glucose	75–120 mg/dL	Normal
	121–180 mg/dL	High (Renal threshold for glucose)
	> 220 mg/dL	Very High
Cholesterol	135–278 mg/dL	Normal
	> 278 mg/dL	High
HbA1c	≤ 3.4	Normal
	> 3.4 to 4.5	Caution zone
	> 4.5 to 8.6	High
Excess urine	Yes / No	–
Excess water intake	Yes / No	–
Urine Glucose	Yes / No	–

2.5 Classification Models

We evaluated several machine classifiers in our

CDDR framework to predict diabetes in canines. Their performance was further evaluated using standard evaluation metrics to identify the most effective classifier for this task. The various machine learning models are as follows:

2.5.1 Random Forest

The algorithm was developed by Breiman in 2001, which is an ensemble-based technique [43]. In it, the model builds many decision trees, where each tree is built using random data samples and random features. The algorithm then selects the most common output from all the trees and has the ability to handle complex relationships between variables and overcome overfitting.

2.5.2 REP Tree

The Reduced Error Pruning Tree (REP Tree) is a fast decision tree model that applies information gain for data split and prunes extra branches so it doesn't overfit [44].

2.5.3 Decision Stump

A simple one-level decision tree classifier that performs predictions with the help of a single attribute. The algorithm serves as a baseline classifier for performance comparison.

2.5.4 LibSVM

There are various implementations of SVM. We applied LibSVM for our experimentation. These machines construct optimal decision boundaries among classes [46]. The algorithm is widely used for classification of complex and non-linear datasets.

2.6 Prediction Process

When a new patient arrives for a clinical checkup, the veterinary care clinic providers initiate a request to the CDDR system. Then the framework retrieves the patient's demographic information, clinical symptoms, and laboratory test results. This information is then integrated with the patient's existing profile and laboratory test results. The combined data, which has the same structure as the training dataset but with an unknown class label, is provided to the trained classifier selected within the CDDR framework.

The trained classifier then predicts whether the patient is diabetic or non-diabetic based on patterns learned during the training phase, as shown in Figure 4. This prediction is subsequently forwarded to a veterinarian for validation.

VetPatientNo	Age	Sex	Breed	Weight	RBS	HbA1c	Lethargy	Excess_Urine	Excess_Water_Intake	Urine_Glucose	Cholesterol	Diagnosed_Class
9987	7	male	German Shephard	34	560	7.6	TRUE	TRUE	TRUE	TRUE	320	?

Figure 4. Sample dataset of an undiagnosed active patient.

2.7 Recommendation Module

The prediction generated by the selected classifier is forwarded to the recommendation module. The diagnosis scale ranges from 1 to 10, where 1–2 indicates an excellent condition, 3–4 a good condition, 5–6 a poor condition, 7–8 a fair condition requiring monitoring, and 9–10 a critical condition [45]. The recommendation module then generates recommendations relating to patient's diet, medication, and healthcare interventions based on the predicted diagnosis level. The validated instance is stored in the dataset and can be incorporated into future retraining of the classifiers to further improve performance and accuracy of CDDR framework.

3 Results

3.1 Experimental Setup

Our experimentations were conducted using the Weka machine learning platform to evaluate the performance of the classification component of the proposed Canine Diabetes Diagnosis and Recommendation (CDDR) framework. Our dataset consists of 181 instances of canine diabetes that were collected from veterinary hospitals, pet care centers. A 10-folds cross-validation was applied to ensure reliable model evaluation. Four machine learning algorithms were evaluated and compared, namely Random Forest (RF), LibSVM, Decision Stump, and REP Tree.

3.2 Classifier Performance

For evaluating the performance of the classifiers various metrics were assessed such as Accuracy, Sensitivity (Recall), Precision, and Mean Absolute Error (MAE). MAE measures the average magnitude of prediction errors, where lower

ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 6, 2026

values indicate better prediction performance. Precision and Recall evaluate the classifier's ability to correctly identify positive cases.
The MAE is defined as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |f_i - y_i| \quad (1)$$

where f_i represents the predicted class and y_i denotes the actual class label. Precision and Recall are calculated as:

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

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

Where *TP*, *FP*, and *FN* represent true positives, false positives, and false negatives, respectively.

The experimental results depict that the Random Forest achieves an accuracy of 93.0%, outperforming all other machine learning classifiers. Furthermore, the classifier produces the highest Precision and Recall of 0.92, which indicates the capability of the classifier in correctly identifying diabetes in canines while minimizing the classification error. The comparative analysis of various metrics relating to LibSVM, Random Forest, Decision Stump, and REP Tree classifiers is shown in the Figure 5, while the MAE comparison is presented in Figure 6. The detailed performance metrics are summarized in Table 2.

Table 2. Performance comparison of tree-based classification models

Classifier	Accuracy (%)	Precision	Recall	MAE
RF	93.0	0.92	0.92	0.07
REP Tree	91.0	0.83	0.83	0.09
LibSVM	82.0	0.44	0.55	0.18
Decision Stump	81.0	0.42	0.63	0.19

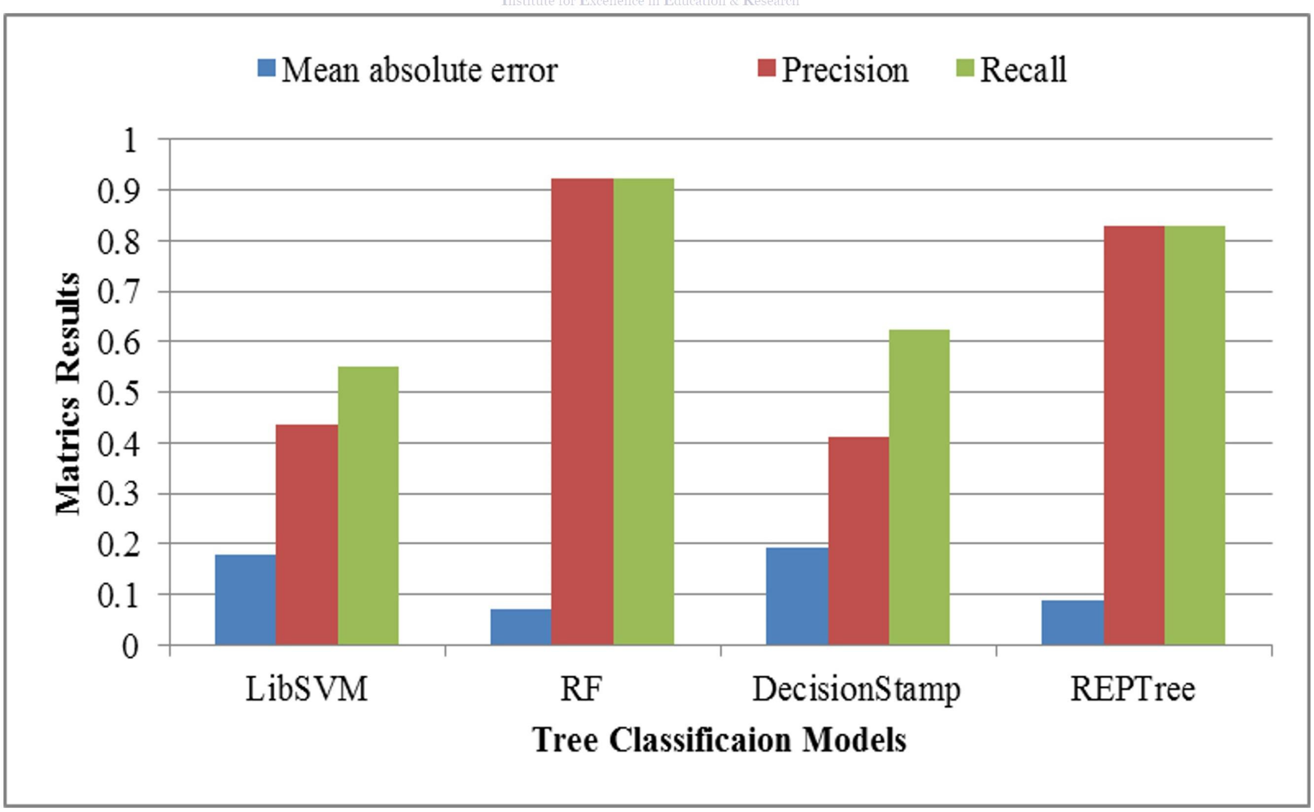

Figure 5. Performance metrics of LibSVM, Random Forest, Decision Stump, and REP Tree classifiers

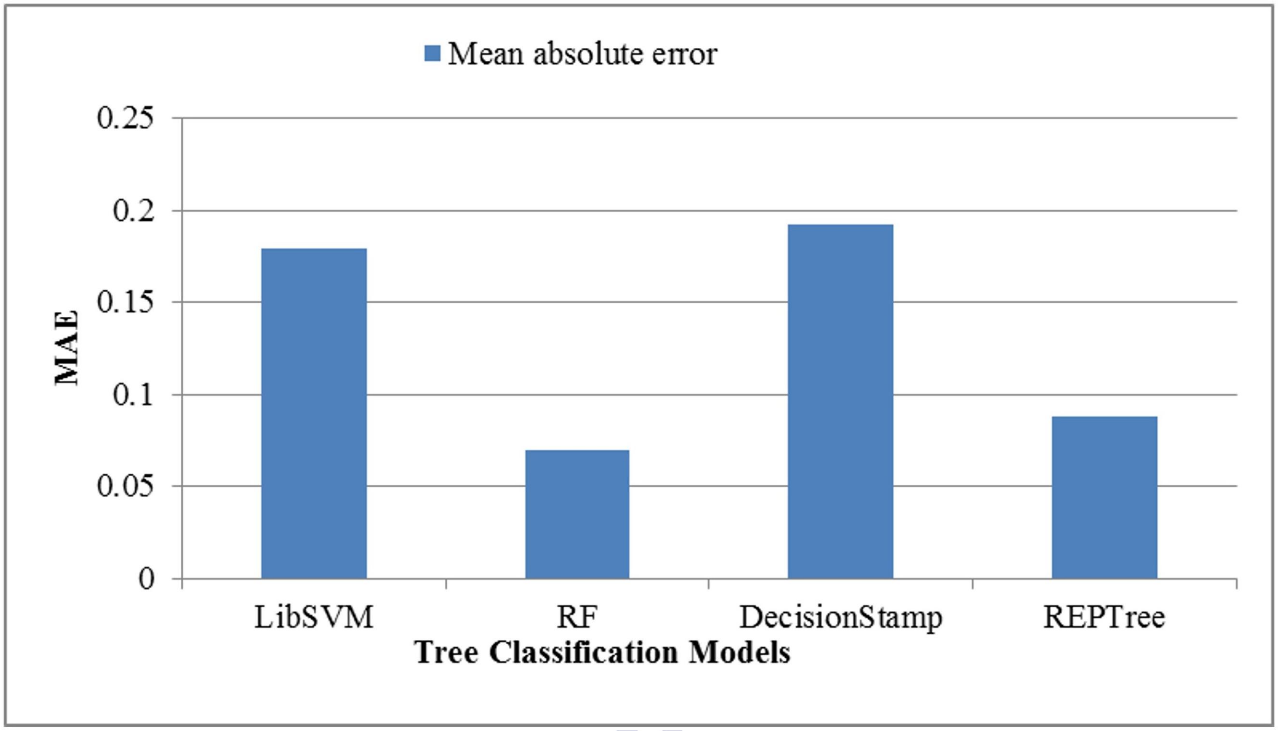

Figure 6. Performance of MAE metric for LibSVM, Random Forest, Decision Stump, and REP Tree

3.3 Comparative Analysis

A comparison was made between various machine learning classifiers, such as, Random Forest, LibSVM, Decision Stump, and the REP Tree. Based on experimental results, Random Forest outperformed other evaluated classifiers across most performance metrics. All classifiers were evaluated using 10-fold cross-validation to ensure a consistent and unbiased comparison. The lower MAE value generated by Random Forest further demonstrates its suitability for canine diabetes prediction. Within our proposed CDDR framework, Random Forest achieved the highest predictive performance, providing a low error rate and high classification accuracy for canine diabetes prediction.

3.4 Framework Output Validation

To assess the practical applicability of the CDDR framework, the prediction outputs and generated recommendations were reviewed by veterinary practitioners. The recommendations relating to diet, medication, and clinical management were found to be consistent with established veterinary practices. The validation process confirmed that the framework can serve as a decision-support tool for assisting veterinarians in canine diabetes management.

4. Discussions

Machine learning models are widely used across diverse disciplines, including soil classification [47], medical science [23], bioinformatics [24], and agriculture [48-50]. In medical and veterinary domains, these models have demonstrated strong performance in disease prediction and diagnosis, including human diabetes detection and emerging applications in canine diabetes diagnosis and management [28,32-34, 51].

The experimental results indicate that the Random Forest classifier selected within the proposed CDDR framework provides effective prediction of canine diabetes. According to Table 2, Random Forest has achieved the highest accuracy of 93.0% while providing a precision and recall of 0.92. It also has the lowest MAE of 0.07 among all the selected classifiers. Several characteristics of the Random Forest classifier contribute to its superior predictive performance.

First, Random Forest effectively captures relationships between clinical and laboratory variables that are nonlinear and complex. Diabetes is usually diagnosed by interaction of more than one physical measure such as sugar level, body weight, age, insulin etc. These complex relationships can be modeled easily by using Random Forest making it well-suited for veterinary medical data.

Secondly, the ensemble nature of Random Forest makes it resistant to noise and overfitting. By integrating predictions from many decision trees, the variance of overall performance reduces. Having this feature is particularly important when using relatively small veterinary datasets, like this study's 181-record dataset.

Finally, the Random Forest can deal with mixed attributes types, numerical and categorical types frequently seen in veterinary medical records.

These findings are supported by previous diabetes prediction studies. As per the report of Maniruzzaman et al. [52], Random Forest provides better diagnostic performance due to its ensemble learning approach as compared to individual classifiers. Similar assessments have been made in studies predicting diabetes in humans, in which Random Forest and various ensemble techniques outperform traditional techniques like SVM and decision tree due to their ability to learn complex patterns, thereby preventing overfitting [53,54].

The CDDR framework offers several benefits for practitioners. They can use the system to assess dogs' diabetes risk in a rapid and consistent manner. Decision-support capabilities are especially beneficial in remote locations, where clinical resources are limited and access to veterinary doctors may be limited. Recommendation component of the framework can specially help in making appropriate decisions based on the predicted severity of the patient. The CDDR framework has the potential to facilitate timely intervention, treatment planning, and improved disease management, particularly for chronic diabetic cases. By integrating prediction and recommendation capabilities, the proposed CDDR framework has the potential to enhance veterinary decision making and improve clinical outcomes for diabetic dogs.

5. Conclusion and Future Work

Recommender systems and machine learning models can greatly assist in disease diagnosis in canines. Our machine learning based Canine Diabetes Diagnosis and Recommendation (CDDR) framework can assist veterinarians in the early detection and management of diabetes in canines. The framework integrates clinical, laboratory, and demographic data collected from veterinary hospitals and pet care centers. Feature selection was performed using an Information Gain method. The CDDR framework generates predictions and treatment recommendations that are vetted by veterinary professionals. Among the evaluated machine learning classifiers, Random Forest achieved the highest performance, with an accuracy of 93.0%, precision and recall of 0.92, and the lowest MAE of 0.07. These results demonstrate that the classifier not only capture complex nonlinear relationships between clinical indicators and symptoms, while also handles overfitting effectively due to its ensemble nature, thereby outperforming LibSVM and Decision Stump particularly in stability and generalization. The proposed CDDR framework can serve as an effective decision-support tool, enabling veterinarians to make timely, accurate, and reliable clinical decisions, thereby significantly reducing risk of diabetes in veterinary patients. In future, deep learning algorithm can be used to enhance the detection accuracy and reliability of the model.

References

1. Najafipour, H. et al. Prevalence and incidence rate of diabetes, pre-diabetes, uncontrolled diabetes, and their predictors in the adult population in southeastern iran: findings from kercadr study. Frontiers in public health 9, 611652 (2021).
2. Standl, E., Khunti, K., Hansen, T. B. & Schnell, O. The global epidemics of diabetes in the 21st century: Current situation and perspectives. European journal of preventive cardiology 26,

7–14 (2019).
3. Otegenova, A., Kazbekova, A., Kulzhanov, M. & Akanov, Z. Navigating the global challenge of diabetes mellitus: insights from kazakhstan's healthcare landscape and strategies for improved management. Interdisciplinary Approaches to Medicine 5, 30–39 (2024).
4. Abualhasan, M., Awad, M. & Sweileh, W. Drug utilization pattern among type II diabetic patients in Palestine. BMC Health Services Research 25, 891 (2025).
5. Baker, E. A. & Fortin, P. T. Introduction, demographics, and epidemiology of diabetes. In The Surgical Management of the Diabetic Foot and Ankle, 1–7 (Springer, 2016).
6. Romero, L. F. Diabetes: the current state of affairs from a population management view. MLO: Medical Laboratory Observer 48, 12–20 (2016).
7. McNally, R. T. Improving Pre-Diabetes Knowledge and Management Among Adults in Primary Care Using Text Messaging (Salve Regina University, 2020).
8. Yudkin, J. S. & Montori, V. M. The epidemic of pre-diabetes: the medicine and the politics. Bmj 349 (2014).
9. Rowley, W. R., Bezold, C., Arikan, Y., Byrne, E. & Krohe, S. Diabetes 2030: insights from yesterday, today, and future trends. Population health management 20, 6–12 (2017).
10. Afrin, N. Risk Factors of Diabetes Mellitus; a Review. Ph.D. thesis, East West University (2016).
11. Ali, M. K., Galaviz, K. I., Weber, M. B. & Narayan, K. V. The global burden of diabetes. Textbook of diabetes 65–83 (2017).
12. Ogurtsova, K. et al. Idf diabetes atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes research and clinical practice 128, 40–50 (2017).
13. Liu, Q., He, X. & Liu, Y. Burden of type 2 diabetes in working-age adults (20–54 years): a gbd 2021 analysis projecting trends to 2035 and exploring the potential benefits of physical activity. Frontiers in Public Health 13, 1706523 (2025).
14. Fleeman, L. & Rand, J. Beyond insulin therapy: achieving optimal control in diabetic dogs. Waltham Focus 15, 12–19 (2005).
15. Nelson, R. W. & Reusch, C. E. Classification and etiology of diabetes in dogs and cats. J Endocrinol 222, T1–T9 (2014).
16. Ringstad, N., Lingaas, F. & Thoresen, S. Breed distributions for diabetes mellitus and hypothyroidism in norwegian dogs. Canine medicine and genetics 9, 9 (2022).
17. Heeley, A. M. et al. Diabetes mellitus in dogs attending uk primary-care practices: frequency, risk factors and survival. Canine Medicine and Genetics 7, 6 (2020).
18. Fall, T. et al. Diabetes mellitus in elkhounds is associated with diestrus and pregnancy. Journal of Veterinary Internal Medicine 24, 1322–1328 (2010).
19. Konishi, K. et al. Increase in the prevalence of canine diabetes mellitus in japan from 2015 to 2023: insights from insurance and clinical data. Journal of Veterinary Medical Science 88, 389–393 (2026).
20. Greco, D. S. Diabetes mellitus in animals: diagnosis and treatment of diabetes mellitus in dogs and cats. In Nutritional and therapeutic interventions for diabetes and metabolic syndrome, 507–517 (Elsevier, 2018).
21. Heeley, A. M., Brodbelt, D. C., O'Neill, D. G., Church, D. B. & Davison, L. J. Assessment of glucocorticoid and antibiotic exposure as risk factors for diabetes mellitus in selected dog breeds attending uk primary-care clinics. Veterinary Record 192, no–no (2023).
22. Ahmed, L., Iqbal, M. M., Aldabbas, H., Khalid, S., Saleem, Y., & Saeed, S. (2023). Images data practices for semantic segmentation of breast cancer using deep neural network. Journal of Ambient Intelligence and Humanized Computing, 14(11), 15227-15243.
23. Ali, A. S., Iqbal, M. M., Khan, A. H., Hameed, N., & Bibi, S. (2023). Lung cancer detection using convolutional neural networks from computed tomography images. Journal of Computing & Biomedical Informatics, 6(01), 133-143.
24. Olson, R. S., La Cava, W., Mustahsan, Z., Varik, A. & Moore, J. H. Data-driven advice

for applying machine learning to bioinformatics problems. arXi preprint arXiv:1708.05070 (2017).
25. Jiang, X., Hu, Z., Wang, S. & Zhang, Y. Deep learning for medical image-based cancer diagnosis. Cancers 15, 3608 (2023).
26. Sahu, H. et al. Analysis of deep learning methods for healthcare sector-medical imaging disease detection. Contemporary Mathematics 830–852 (2023).
27. Al-Alshaikh, H. A. et al. Comprehensive evaluation and performance analysis of machine learning in heart disease prediction. Scientific Reports 14, 7819 (2024).
28. Chang, V., Ganatra, M. A., Hall, K., Golightly, L. & Xu, Q. A. An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators. Healthcare Analytics 2, 100118 (2022).
29. Karunakaran, D. & Chandran, R. K. Deep learning based diabetes mellitus prediction for healthcare monitoring. Journal of Electrical Engineering & Technology 18, 4399–4413 (2023).
30. Doguc, O., Bilgi, S. B., Cagdas, S. & Yilmazturk, N. A decision support system for detecting fip disease in cats based on machine learning methods. In International Conference on Emerging Trends and Applications in Artificial Intelligence, 176–186 (Springer, 2023).
31. Schofield, I. et al. Machine-learning based prediction of cushing's syndrome in dogs attending uk primary-care veterinary practice. Scientific Reports 11, 9035 (2021).
32. Razaq, A., Ramzan, S., Jabbar, S., Iqbal, M. M., Habib, M. A., & Raza, U. (2025). A Deep Learning Framework for Early Parkinson's Disease Detection: Leveraging Spiral and Wave Handwriting Tasks with EfficientNetV2-S. Diagnostics, 15(21), 2795.
33. Kour, S. et al. Artificial intelligence and its application in animal disease diagnosis. Journal of Animal Research 12, 1–10 (2022).
34. Chafai, N., Hayah, I., Houaga, I. & Badaoui, B. A review of machine learning models applied to genomic prediction in animal breeding. Frontiers in genetics 14, 1150596 (2023).
35. Nayak, S., Jena, L., Palai, P., Mishra, S. & Swain, M. K. Application of machine learning in the analysis and prediction of animal disease. In Sustainable Farming Through Machine Learning, 207–219 (CRC Press, 2024).
36. Javeed, M., Aslam, S., Farhan, M., Aslam, M., & Khan, M. (2023). An Enhanced Predictive Model for Heart Disease Diagnoses Using Machine Learning Algorithms. Technical Journal, 28 (04), 64–73.
37. Zhou, X. et al. The early prediction of common disorders in dairy cows monitored by automatic systems with machine learning algorithms. Animals 12, 1251 (2022).
38. Uddin, M. M. et al. Magnitudes of diseases in dogs vary among different levels of age, gender, breed, and season: A hospital-based, retrospective cross-sectional study. Heliyon 7 (2021).
39. Buvik, R. Genetics of endocrine diseases in Miniature Schnauzer:(Review of literature). Ph.D. thesis (2014).
40. Singh, S. STUDY ON DIABETES MELLITUS AND ITS MANAGEMENT IN DOGS. Ph.D. thesis, Kashmir University (2022).
41. Norris, O. & Schermerhorn, T. Relationship between hba1c, fructosamine and clinical assessment of glycemic control in dogs. PLoS One 17, e0264275 (2022).
42. Teixeira, F. A. & Brunetto, M. A. Nutritional factors related to glucose and lipid modulation in diabetic dogs: literature review. Brazilian Journal of Veterinary Research and Animal Science 54, 330–341 (2017).
43. Salman, H. A., Kalakech, A. & Steiti, A. Random Forest algorithm overview. Babylonian Journal of Machine Learning 2024, 69–79 (2024).
44. Mohamed, W. N. H. W., Salleh, M. N. M. & Omar, A. H. A comparative study of reduced error pruning method in decision tree algorithms. In 2012 IEEE International conference on control system, computing and engineering, 392–397 (IEEE,2012).
45. Iba, W. & Langley, P. Induction of one-level

decision trees. In Machine learning proceedings 1992, 233–240 (Elsevier, 1992). 11/12
46. Abdullah, D. M. & Abdulazeez, A. M. Machine learning applications based on svm classification a review. Qubahan Academic Journal 1, 81–90 (2021).
47. Heung, B. et al. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. Geoderma 265, 62–77 (2016).
48. Liakos, K., Busato, P., Moshou, D., Pearson, S. & Bochtis, D. Machine learning in agriculture: A review. Sensors 18, 2674 (2018).
49. Chlingaryan, A., Sukkarieh, S. & Whelan, B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and electronics in agriculture 151, 61–69 (2018).
50. Raffini, F. et al. From nucleotides to satellite imagery: Approaches to identify and manage the invasive pathogen xylella fastidiosa and its insect vectors in europe. Sustainability 12, 4508 (2020).
51. Sholehrasa, H. Leveraging predictive modeling and explainable AI to understand veterinary safety profiles and health outcomes using OpenFDA data. Ph.D. thesis (2025).
52. Maniruzzaman, M. et al. Accurate diabetes risk stratification using machine learning: role of missing value and outliers. Journal of medical systems 42, 92 (2018).
53. Oehm, A. W. et al. Random Forest classification as a tool in epidemiological modelling: Identification of farm-specific characteristics relevant for the occurrence of fasciola hepatica on german dairy farms. Plos one 18, e0296093 (2023).
54. Khan, A. A., Chaudhari, O. & Chandra, R. A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. Expert Systems with Applications 244, 122778 (2024).