ISSN NO: 9726-001X

Volume 13 Issue 02 2025

JOURNAL OF CURRENT SCIENCE

Proactive Live Stone Care: Leveraging Machine Learning for Disease Prediction and Classification

[1]SHAIK FIRDOS, [2] MEDA OME SAI LAKSHMI SANTHOSHI, [3] NADIPALLI SINDHU, [4] BESE BHANU PRAKASH VARMA, [5]Mrs. N. BHARATHI,

[1234]Student Department of CSE, DNR College of Engineering & Technology, Balusumudi, Bhimavaram, India.

[5]Assistant Professor, Department of CSE, DNR College of Engineering & Technology, Balusumudi, Bhimavaram, India.

Abstract –

Stock animals are those that are raised for the purpose of making a profit. Most livestock animals reside in remote areas where they don't have easy access to veterinarians. Generally speaking, it is not an easy process to diagnose a condition quickly and accurately. Identifying the illness's result and taking suitable safeguards is vital to avoid the sickness from spreading among the cattle. So, it's necessary to have a system that can help predict cattle illnesses from their signs and then suggest ways to avoid such diseases. Here, we apply eight different machine learning techniques to a dataset consisting of farmed animal health concerns detected by automated monitoring systems. Age, symptoms, temperature fluctuations, and illness kind are the factors that are examined by the machine learning algorithms. We have used machine learning techniques to a dataset with the goal of early illness prediction and diagnosis in cattle. Integration of feature selection methods with machine learning techniques such as logistic regression, XGBoost, Catboost, light gradient boost, Random Forests, and Support Vector Machines allowed for the prediction of illnesses. After looking at all of the data, the Random Forest classifiers were the most dependable and had the highest accuracy rate at 100%. Ensemble learning, machine learning, illness diagnosis, classification algorithm, and animal disease prediction are all terms that might be used as keywords.

INTRODUCTION

The global spread of cattle illness has a devastating impact on domestic animals and leads to a dramatic rise in animal fatality rates [1]. We are living in a digital era of high-throughput in cattle management. Every day and hour, using massive volumes of data gathered by systems of biosensors, acoustical, mechanical, and electro-optical sensors, choices are made based on quantitative and qualitative analytical findings [2]. Serious diseases in animals pose a serious risk to both human and animal health. The importance of big data analysis and illness prediction in animals is growing as a result of the abundance of data accessible and the rapid pace of globalization [3]. Most sub-clinical illnesses go undetected until they cause noticeable clinical signs because farmers aren't aware of how to avoid and identify them early. Institution of Computer Science and Engineering, G. Prabu Kanna School of VIT Bhopal University, Central India Attn: gpkanna | Gmail: [4]. Because of this, the negative impacts on dairy cows' health and output last longer, and therapy is more difficult and expensive [2]. Consequently, it stands to reason that pre-diagnosis data that can more precisely identify a disease or forecast an animal's risk would be more useful for early detection and intervention than clinical symptoms. Machine Learning (ML) is a subfield of AI that uses large datasets and statistical approaches to predict when cattle will become sick or how well they will perform [5, 6]. Learning from data and applying that knowledge to analysis and prediction is the goal of ML algorithms [4, 7, 8]. For the purpose of early investigation of illness prediction and categorization in cattle, this research used ML algorithms. Researchers have used ML approaches such as logistic generalized linear mixed models, decision-tree induction, neural networks, eXtreme Gradient boosting, naïve Bayes, and random forests to detect or predict various health issues, including clinical mastitis [9], [10]. Heart disease, metabolic syndrome in humans, and animal illnesses are only a few examples of the many other conditions that may have very accurate prognostic and diagnostic models developed via the effective use of diverse anthropometric and medical data that is easily accessible [7], [11]-[13]. Recent advances in data processing, especially in the fields of artificial intelligence (AI), machine learning (ML), and healthcare, have made this a reality. With its robust data analysis capabilities, ML is an

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

JOURNAL OF CURRENT SCIENCE

algorithm-based data analysis tool that can handle complex interdependencies between variables [14]. Considering these characteristics, a model for risk prediction based on machine learning was developed for the purpose of this experiment in order to diagnose diseases in cattle. Several ML classifiers, including Random Forests (RF), were used and studied for their efficacy in illness prediction and classification in this study.

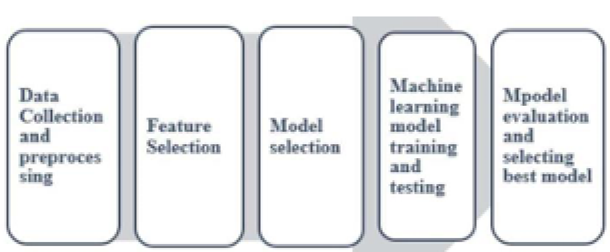
RELATED STUDIES

In the realm of agriculture, ML algorithms were used for the purpose of predicting and diagnosing cow diseases. More recently, ML has been used to study the patterns of mastitis pathogen transmission in cattle and to diagnose preclinical and clinical mastitis in individual animals [9], [15], [16]. The authors of [17] have found accelerometer data attached to cows and created an algorithm to detect lameness using this information. In order to detect mastitis and foot and mouth disease, the authors of [18] proposed a technique that makes use of an Internet of Things infrastructure. In their system, the authors have taken many things into account, such as a microcontroller, a machine learning algorithm, and other components like motion, sound, temperature, and so on. In order to detect breeding trends in cows, such as the beginning, peak, and end of estrus, a system is described in [19] that analyses three-axis acceleration data from Internet of Things sensors. The system examines data on three-dimensional acceleration retrieved from Internet of Things sensors. Research on dairy calves has shown that some illnesses may be predicted using logistic regression and gradient boosting ML algorithms. These disorders include anestrus, hyperketonemia, lameness, metritis, and ovarian cysts. The findings point to the complex interplay between climate, nutrition, and housing circumstances as the root cause of disease in dairy cows [20]. The efficacy of several ML models in disease resistance prediction was assessed in the research in [21] using both real and virtual datasets. In order to examine the efficacy of disease prediction in livestock and aquaculture species, the following models were utilized: DT, SVM, RF, XGBoost, Adaboost, and the genomic best linear unbiased prediction for threshold characteristics backed by Markov chain Monte Carlo. To anticipate instances of lameness in dairy cows, ML models have been used, for example, by using conformation and milk production parameters [22].

While it is true that the classifiers used in this research were not very effective, expanding the data set for model training might lead to better prediction performance. The results show that machine learning algorithms may be used to a wide variety of disease factors to accurately forecast when a complicated illness, such as lameness in dairy cows, would manifest. Animal health monitoring is seeing a growing usage of ML algorithms for syndromic surveillance data processing and extraction from clinical records [23]. Recent studies have shown that using machine learning algorithms to automate the mining of free-text data in clinical and post-mortem reports may greatly assist with the adoption of animal health syndromic monitoring [24], [25]. With the help of precision technology, farmers and veterinarians can collect and analyze massive amounts of data from their farms, which may greatly enhance health and production management. This kind of data can be analyzed using machine learning algorithms, much like text data, which may make syndrome monitoring easier. In their study, the authors examined ML models that might predict the likelihood of a contagious or inflammatory central nervous system illness in impaired cattle [26-29]. The efficacy of six distinct ML algorithms (LR, SVM, RF, KNN, MLP, GB) to predict the presence of infectious or inflammatory illness was compared using demographics, findings from neurological tests, and analysis of cerebrospinal fluid (CSF). The RF model outperformed all of the others in terms of accuracy, which was 100%. Here, we compare six machine learning classifiers—including statistical, boosting, and bagging models—using a cattle dataset to their prior *work.*

METHODOLOGY

Figure 1 shows the technique, which details the machine learning algorithm used to predict cattle data.



ISSN NO: 9726-001X

Volume 13 Issue 02 2025

Figure 1. Flowchart for livestock disease prediction

The proposed setup tests six different ML classifiers: SVM, RF, KNN, LGB, CatBoost, and XGBoost. A dataset of farm animals pulled from the Kaggle repository is used to evaluate the classifiers. Nevertheless, null values and other missing values were removed using data pretreatment methods. We have picked the relevant features from the dataset and deleted the extraneous ones using the feature selection approach to boost the learning accuracy. Figure 2 displayed the relationship between the important characteristics used for training the ML model. Figures 3–5 displayed a synopsis of the dataset characteristics used for classification.

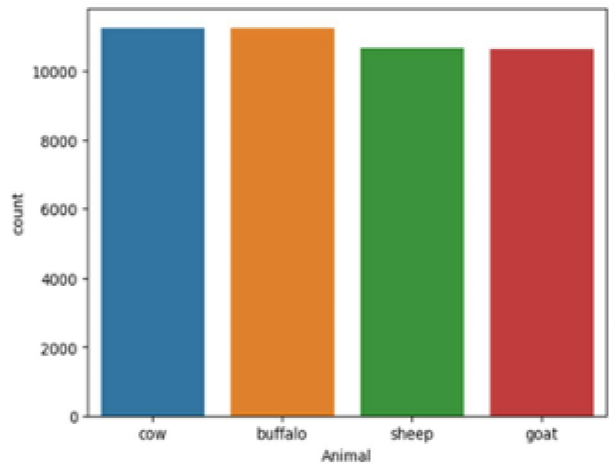


Figure 2. Data classification: Types of livestock animals

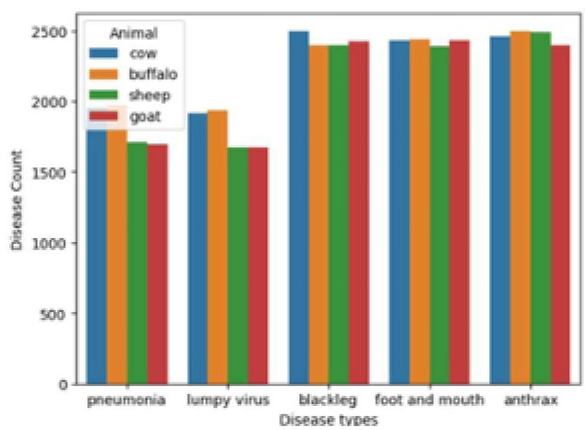


Figure 3. Disease types in animals

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

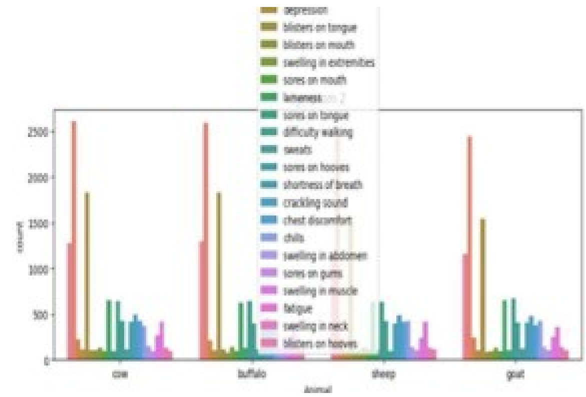


Figure 4. Symptoms type-1 for animal disease

RESULTS

The results of applying machine learning methods to the dataset of animal diseases are detailed in this section. We selected six separate machine learning models for animal illness prediction based on their proficiency with categorical datasets. We then employed F-1 scores, recall, accuracy, precision, confusion matrices, and loss to assess the models. Each classifier's confusion matrix is shown in Figures 6–11 in its own section: RF, KNN, SVM, LGB, XGBoost, and Catboost. A. RF was RFs classify using a bagging ensemble. Using the training data, RFs generate a large number of decision trees. For each data subset, a decision tree set is built to provide a classification result; however, from the whole set, only a minimum data sample is used.

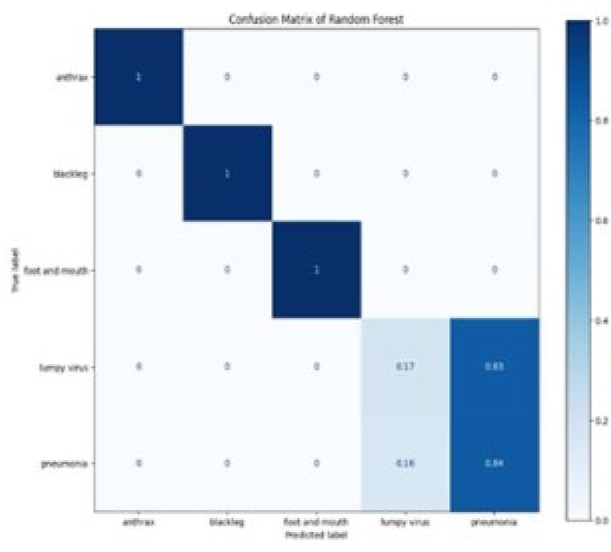


Figure 7. Confusion Matrix: Knn

Part B. KNN The k nearest neighbors method finds the average or most frequent attribute across a collection by grouping together the items that are geographically closest to each other. Using eq. 1, the KNN approach determines the distance between the closest spots.

$$d(a, b) = \sum_{k=1}^{n} a_i - b_i \qquad (1)$$

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

Simple Vs. Finding the best decision boundary in a high-dimensional space is mostly done by an SVM using its input characteristics. Following its establishment, the classifier takes use of this boundary. The main goal of support vector machines (SVMs) is to choose a hyperplane from a huge set of options using the purpose of maximizing the margins over all data samples.

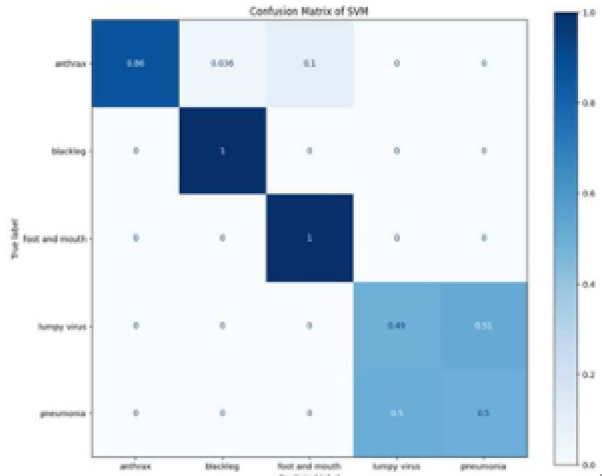


Figure 8. Confusion Matrix:

Support Vector Machine D. LGB LightGBM is a decision-tree based gradient-boosting framework that reduces memory use and improves model efficiency. All gradient boostin algorithms are based on the histogram-based approach, although these alternatives address its limitations.

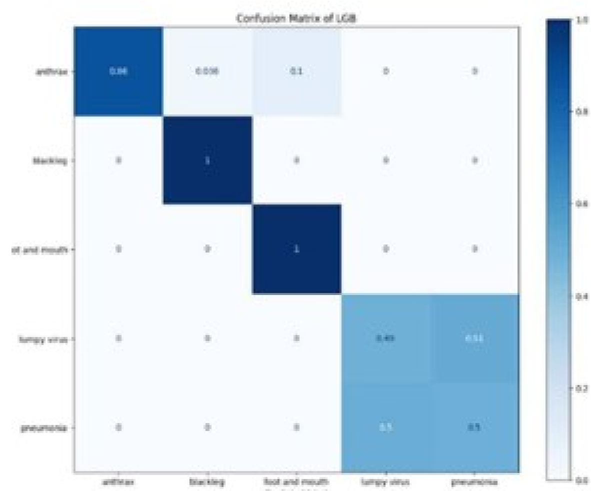


Figure 9. Confusion Matrix:

The boosting approach known as LGB E. XGBoost Gradient Boosting expands upon the stagewise addition method. The strong learner algorithm is the result of combining many weak learning algorithms that were trained on the same dataset. When it comes to categorical datasets, LGB really shines, and the binning or bucketing method is what it uses to manage the category features. For the purpose of automating data handling, we have transformed every LGB characteristic into a category datatype. This procedure involves computing the variance of every data sample

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

and then sorting the results by decreasing variance. Because of their shown excellent performance, low variance data samples will be assigned less weight throughout the dataset sampling process.
F. Catboost When it comes to categorical datasets, the CatBoost model outperforms conventional machine learning techniques. Based on the target variable, CatBoost encodes its categorical features. To further increase the model's accuracy, the target variable is taken into account d

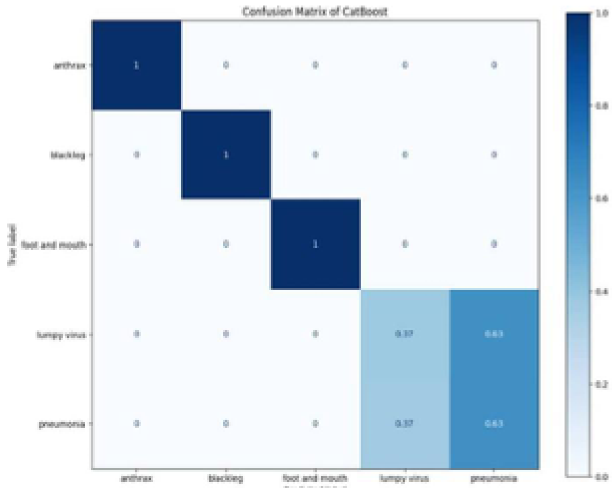


Figure 11. Confusion Matrix:

Table 1 below shows the results of each classifier in terms of accuracy, precision, recall, F1-score, and loss when it comes to predicting cattle diseases using Catboost. Based on these results, it seems that the Random Forest and catboost classifiers are the best options for predicting cattle diseases from categorical datasets.

TABLE I. PERFORMANCE COMPARISON RESULTS OF DIFFERENT ML MODELS FOR LIVESTOCK DISEASE PREDICTION

Model	Weighted Average				
	Accuracy	Precision	Recall	F1-score	Loss
RF	**100**	**0.84**	**0.84**	**0.82**	**0.16**
KNN	82.32	0.80	0.80	0.80	0.16
SVM	80.37	0.81	0.80	0.80	0.19
LGB	80.37	0.81	0.80	0.80	0.19
XGBoost	82.49	0.82	0.82	0.82	0.17
Catboost	**83.72**	**0.84**	**0.84**	**0.83**	**0.16**

CONCLUSION

From hobby farmers raising a few animals to large-scale cattle breeders and processors, the livestock sector is vital to the country's economy. The profitability of any cattle operation is directly affected by the health of the animals. Early detection and prevention are crucial since diseases worsen and kill more rapidly as they develop. This research used a range of machine learning models with 10-fold cross-validation, such as kNN, SVM, RF, XGBoost, stacking, and bagging, to predict the likelihood of illnesses in agricultural animals. The best model for predicting the probability of a cattle animal's disease diagnosis from a variety of symptoms was ultimately determined by comparing and evaluating the models using attributes like recall, accuracy, precision, and F1 score, among others.

The Random Forest model had the best results when tested with 10-fold cross-validation; it achieved 100% accuracy, with precision and recall values of 0.84, and an F1 score of 0.83.

REFERENCES

[1]. D. Ashar, A. Kanojia, R. Parihar, and S. Kudoo, "Livestock Disease Prediction System," Int. J. Res. Innov., vol. 1, no. 4, pp. 2581–7280, 2021.

[2]. T. Hollings, A. Robinson, M. V. Van Andel, C. Jewell, and M. Burgman, "Species distribution models: A comparison of statistical approaches for livestock and disease epidemics," PLoS One, vol. 12, no. 8, 2017, doi: 10.1371/journal.pone.0183626.

[3]. G. Franzo, M. Legnardi, G. Faustini, C. M. Tucciarone, and M. Cecchinato, "When Everything Becomes Bigger: Big Data for Big Poultry Production," Animals, vol. 13, no. 11. 2023. doi: 10.3390/ani13111804.

[4]. K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," Sensors (Switzerland), vol. 18, no. 8. 2018. doi: 10.3390/s18082674.

[5]. J. K. Gwaka et al., "Machine-Learning Approach for Risk Estimation and Risk Prediction of the Effect of Climate on Bovine Respiratory Disease," Mathematics, vol. 11, no. 6, 2023, doi: 10.3390/math11061354.

[6]. P. Tripathi, N. Kumar, M. Rai, P. K. Shukla, and K. N. Verma, "Applications of Machine Learning in Agriculture," in Smart Village Infrastructure and Sustainable Rural Communities, 2023. doi: 10.4018/978-1-6684-6418-2.ch006.

[7]. J. Lasser et al., "Integrating diverse data sources to predict disease risk in dairy cattle-a machine learning approach," J. Anim. Sci., vol. 99, no. 11, 2021, doi: 10.1093/jas/skab294.

[8]. D. D. Martinelli, "Machine learning for metabolomics research in drug discovery," Intell. Med., vol. 8, 2023, doi: 10.1016/j.ibmed.2023.100101.

[9]. M. Ebrahimi, M. Mohammadi-Dehcheshmeh, E. Ebrahimie, and K. R. Petrovski, "Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: Deep Learning and Gradient- Boosted Trees outperform other models," Comput. Biol. Med., vol. 114, 2019, doi: 10.1016/j.compbiomed.2019.103456.

[10]. N. Modi and J. Singh, "A comparative analysis of deep learning algorithms in eye gaze estimation," 2022. doi: 10.1109/ICDABI56818.2022.10041480.

[11]. G. Zou, Q. Zhong, P. OUYang, X. Li, X. Lai, and H. Zhang, "Predictive analysis of metabolic syndrome based on 5-years continuous physical examination data," Sci. Rep., vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-35604-8.

[12]. H. Shin, S. Shim, and S. Oh, "Machine learning-based predictive model for prevention of metabolic syndrome," PLoS One, vol. 18, no. 6 June, 2023, doi: 10.1371/journal.pone.0286635.

[13]. V. Soares de Siqueira, M. Marcos Borges, R. Gomes Furtado, C. Nunes Dourado, and R. Martins da Costa, "Artificial intelligence applied to support medical decisions for the automatic analysis of echocardiogram images: A systematic review," Elsevier, vol. 120, p. 102165, 2021, doi: 10.1016/j.artmed.2021.102165.

[14]. M. Saberi-Karimian et al., "Potential value and impact of data mining and machine learning in clinical diagnostics," Crit. Rev. Clin. Lab. Sci., vol. 58, no. 4, pp. 275–296, 2021, doi: 10.1080/10408363.2020.1857681.

[15]. R. M. Hyde et al., "Automated prediction of mastitis infection patterns in dairy herds using machine learning," Sci. Rep., vol. 10, no. 1, 2020, doi: 10.1038/s41598-020-61126-8.

[16]. P. S. Chatterjee, N. K. Ray, and S. P. Mohanty, "LiveCare: An IoT- Based Healthcare Framework for Livestock in Smart Agriculture," IEEE Trans. Consum. Electron., vol. 67, no. 4, 2021, doi: 10.1109/TCE.2021.3128236.

[17]. J. Haladjian, J. Haug, S. Nüske, and B. Bruegge, "A wearable sensor system for lameness detection in dairy cattle," Multimodal Technol. Interact., vol. 2, no. 2, 2018, doi: 10.3390/mti2020027.

[18]. S. Vyas, V. Shukla, and N. Doshi, "FMD and mastitis disease detection in cows using internet of thingh(iot)," in Procedia Computer Science, 2019, vol. 160. doi: 10.1016/j.procs.2019.11.019.

[19]. M. Lee, "IoT Livestock Estrus Monitoring System based on Machine Learning," Asia-pacific J. Converg. Res. Interchang., vol. 4, no. 3, 2018, doi: 10.21742/apjcri.2018.09.12.

[20]. J. Lasser et al., "Integrating diverse data sources to predict disease risk in dairy cattle-a machine learning approach," J. Anim. Sci., vol. 99, no. 11, Nov. 2021, doi: 10.1093/jas/skab294.

[21]. C. Palaiokostas, “Predicting for disease resistance in aquaculture species using machine learning models,” Aquac. Reports, vol. 20, 2021, doi: 10.1016/j.aqrep.2021.100660.