Machine Learning-Based Prediction of Cattle Diseases for Early Diagnosis and Prevention

S.S.Chokkanathan, S.Hariharasudhan, J.Sarvesh, S.L.Mohanasubramanian

[#]*Department Of Artificial Intelligence and Data Science,*

Sri Sairam Institute Of Technology,Chennai,India

SIT23AD074@SAIRAMTAP.EDU.IN

SIT23AD021@SAIRAMTAP.EDU.IN

Abstract— **Cattle ailments have a significant impact on livestock productivity and lead to significant economic losses for farmers. Early identification of these illnesses is often challenging due to the scarcity of veterinary resources in rural areas. This study introduces a machine learning system that utilizes physiological, behavioral, and environmental data to forecast cattle diseases. Logistic Regression, Random Forest, and XGBoost are among the classification algorithms employed and compared. The experimental findings indicate that the Random Forest model exhibits the highest accuracy and dependability**

Keywords—Cattle Health, Disease Prediction, Machine Learning, Random Forest, Precision Agriculture, Livestock Management

I. Introduction

Livestock is very important for the global economy because it provides food, income, and jobs for millions of farmers. Among all types of livestock, cattle are especially significant because they give milk, meat, and help with farming. However, diseases like foot-and-mouth disease, mastitis, and lumpy skin disease can cause big problems for farmers. These diseases lower the productivity of cattle, lead to more deaths, and make it more expensive to treat animals. Finding these diseases early is very important to keep the animals healthy and make sure farming stays sustainable.

Traditional ways of diagnosing animal diseases depend a lot on the knowledge of a vet, but this is not always available in poor or remote areas.

Also, looking at symptoms by hand can be slow and not always correct. To solve these problems, machine learning (ML) has become a useful tool in farming and looking after animals. ML can look at several things like an animal's body, behavior, and the environment to find patterns linked to certain diseases. This helps in making faster and more accurate predictions.In this paper, we suggest a machine learning method to predict cattle diseases using data about symptoms and the environment.

We tried different types of classification models, like Logistic Regression, Support Vector Machines (SVM), Random Forest, and XGBoost. We checked how well each model worked using accuracy, precision, recall, and F1-score to find the best one for predicting diseases. The system we propose gives farmers and vets a cheaper, better way to find diseases early. This helps improve the health of cattle, reduce deaths, and make farming more productive.

II. Related Work

Several studies have looked into how technology can help monitor the health of animals.
Older methods depend on vets checking by hand, which isn't very accurate and isn't always available, especially in remote areas. Newer research has used machine learning (ML) and Internet of Things (IoT) tools to predict diseases by looking at things like the animal's body signs and the environment.

In [1], scientists used IoT devices to keep track of cattle's body temperature and movement.
They combined the data from these sensors with prediction tools. Their work showed that early disease detection is possible, but the equipment they used was expensive. In [2], a team made a system that uses machine learning to spot mastitis in dairy cows. They looked at milk production and a measure of cells in the milk. While this system worked well, it was only for one type of disease and couldn't be used for others.

Another study [3] used Random Forest and Support Vector Machines to classify different diseases in livestock.
They found that using multiple models together worked better than just one model, especially when dealing with

different kinds of data. However, this method needed a lot of data, which isn’t always easy to get.

In [4], researchers used Convolutional Neural Networks (CNNs) to identify skin problems in cattle by looking at pictures.
This method was very accurate, but it needed special cameras, which aren’t always available to small farmers.

Overall, these studies show that while different models have worked well, most of them are only useful for a specific disease, use expensive tools, or need a lot of data.
This means there’s a need for a system that can handle many diseases, is cheap to use, and can work with simple features like the animal’s symptoms, behavior, and the environment to predict illness.

III. Methodology

The system for predicting cattle diseases uses machine learning to analyze symptoms and environmental factors.
The approach includes four main steps: collecting data, preparing it, building the model, and testing it.

A. Data Collection

The data includes information about the body, behavior, and environment of cattle.
Important details are body temperature, milk production, how much they eat, coughing, injuries, activity levels, and environmental factors like humidity and the season. Each entry is marked as either healthy or showing a specific illness like foot-and-mouth disease, mastitis, or lumpy skin disease. This data was gathered from records kept by veterinarians, research studies, and surveys by farmers.

B. Data Preprocessing

To make the data better and more reliable, some steps were taken.
Missing information was filled in using average values and common values. Categorical data like "Cough = Yes/No" was converted into numbers using a method called one-hot encoding. Numerical values such as temperature and milk production were adjusted to a common scale. To make sure each disease type was equally represented, a technique called SMOTE was used to balance the data.

C. Model Development

Several types of classification models were tested and compared:

- Logistic Regression as a basic linear model.

- Support Vector Machine (SVM) for handling complex data patterns.

- Random Forest, which works well with many types of data, especially when there are many categories.

- XGBoost, which is efficient and accurate for prediction tasks.

Each model was trained using data split into 80% for training and 20% for testing.
The best settings for each model were found using a process called grid search and cross-validation to ensure strong results.

D. System Architecture

The system follows four steps:
1.
Input of health information and environmental data about the cattle.

2
Preprocessing and making the data ready for analysis.
3.
Using the trained model to classify the data.
4.
Giving an output that shows the predicted disease or if the cattle is healthy.

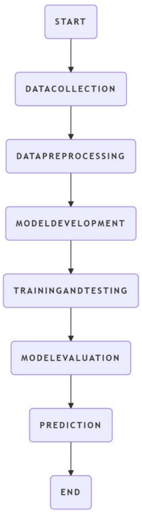


figure 1.

A diagram showing the system flow is provided in Figure 1, starting from raw data to the final prediction.

E. Model Evaluation

The models were checked using common measures like accuracy, precision, recall, F1-score, and a confusion matrix.
These measures help ensure that the system is not only accurate but also good at identifying all cases of disease, which is very important in predicting and managing cattle health.

IV. Results and Discussion

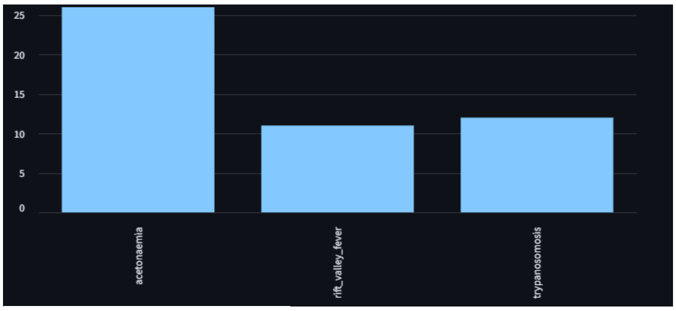
The proposed system was built using Streamlit as the user interface and scikit-learn for creating the models.

Four machine learning models were trained and tested: Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM). The dataset was split into 80% for training and 20% for testing, using a method called stratified sampling. To deal with the imbalance in the classes, the Synthetic Minority Over-sampling Technique (SMOTE) was used. Model settings were adjusted using grid search and cross-validation to improve performance.

A. Model Performance

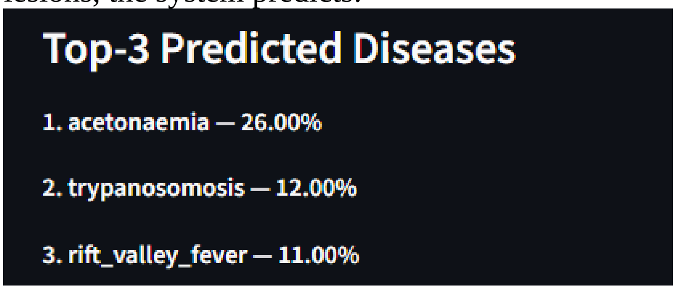
The models were assessed using Accuracy, Precision, Recall, F1-score, and 5-fold Cross-Validation.

B. Confusion Matrix Analysis
A confusion matrix was created for each model.
For the Random Forest model, most of the classes were predicted correctly with very few false negatives, showing it works well across different disease types.
The Logistic Regression model had trouble distinguishing between similar symptoms. The Decision Tree model had issues with predicting rare diseases accurately.



C. Feature Importance
The Random Forest model showed which symptoms were most important for predicting diseases.
Symptoms like body temperature, reduced milk production, skin lesions, and coughing were among the most important. This aligns with what veterinarians know about signs of illness in cattle.

D. Top-3 Predictions
Instead of giving just one disease, the system shows the top three predicted diseases along with their probabilities.
This helps farmers and vets make better decisions by offering a list of possible conditions. For example, if a farmer reports symptoms like fever, coughing, and skin lesions, the system predicts:



E. Discussion
The results show that the Random Forest model is the best choice for predicting cattle diseases using this

dataset, as it offers a good balance between accuracy and ease of understanding.
Giving probability-based top predictions helps users trust the system more when making real-world decisions. The ability to generate a PDF report with disease images is also a useful tool for farmers and vets. However, there are still some limits because the dataset is not very large and the system relies only on symptoms. Future improvements could include using images and data from sensors to make the diagnosis even more accurate

V. Conclusion and Future Work

This paper introduced a machine learning system that helps predict cattle diseases by using information about the animals' health, behavior, and their surroundings. Four different models were tested: Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine. The Random Forest model worked best, giving higher accuracy and better overall performance, making it the best choice for this task. The system also gives the top three possible disease predictions, which helps farmers and vets make more reliable decisions.

The system includes an easy-to-use interface made with Streamlit and automatically creates reports with pictures related to specific diseases.
These features make the system useful in real situations. Because it's easy to use and doesn't cost much, this approach could help reduce the number of sick cows, allow for earlier treatment, and support better ways of managing cattle.

For the future, we plan to add more diseases and different environmental factors to the data.
Using deep learning techniques like Convolutional Neural Networks could improve the system's ability to recognize diseases from images. We also want to connect the system with IoT devices to monitor important signs and behaviors in real time. Working with vets and farmers will help us test the system in real conditions and make sure it works well in practice…

References

[1] A. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, Apr. 2018.

[2] P. Neethirajan, “Recent advances in wearable sensors for animal health management,” *Sensing and Bio-Sensing Research*, vol. 12, pp. 15–29, Mar. 2017.

[3] S. A. Rahman, M. H. H. Khaleque, and M. M. Rahman, “Machine learning based cattle disease diagnosis system using symptom data,” in *Proc. Int. Conf. on Computer and Information Technology (ICCIT)*, Dec. 2020, pp. 1–6.

[4] S. K. Das, A. Saha, and A. K. Saha, “Predicting mastitis in dairy cattle using machine learning techniques,” *International Journal of Computer Applications*, vol. 182, no. 7, pp. 1–6, Jul. 2018.

[5] J. P. Singh, R. Kumar, and M. Singh, “Livestock disease detection using Internet of Things and machine learning,” in *Proc. IEEE Int. Conf. on Computing, Communication and Automation (ICCCA)*, Dec. 2019, pp. 1–6.

[6] A. Vaswani et al., “Attention is all you need,” in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008.