Web-Based Animal Disease Prediction: A Random Forest Approach

Yeshwanth Pilla[1], P. Sanyasi Naidu[2]
[1]*MTech Student, Andhra University College of Engineering (A), Visakhapatnam, Andhra Pradesh*
[2]*Professor, Andhra University College of Engineering (A), Visakhapatnam, Andhra Pradesh*

***Abstract*— The health of livestock is vital for the agricultural economy and the food supply chain. Detecting diseases in animals early can lower mortality rates, protect farmers' incomes, improve the productivity of milk, meat, and fiber, and stop the spread of zoonotic infections. This study presents a user-friendly web-based system for predicting animal diseases. It uses a machine learning model called Random Forest, which is implemented through the Weka library, with a modern full-stack setup. The frontend is created with Angular, while the backend relies on Spring Boot. The dataset contains information about eight animal types and 120 breeds. It includes demographic data, such as age, gender, and weight, physiological data like body temperature and heart rate, temporal data on symptom duration, and 25 binary indicators of symptoms or clinical signs. The trained Random Forest model achieved an impressive 99% accuracy on evaluation data that was set aside. The predicted disease outcomes are returned to the Angular frontend, enabling farmers to quickly book veterinary appointments. This smooth connection between prediction and care improves access to animal healthcare, especially in rural and underserved areas.**

***Keywords:*Animal Disease Prediction, Machine Learning, Random Forest, Weka, Angular, Spring Boot, Veterinary Decision Support, Livestock Health.**

INTRODUCTION

Background & Motivation

Livestock production plays a crucial role in food security, rural income, and national economies around the world. In many regions, such as India, how much dairy, meat, and fiber can be produced hinges on quickly detecting and treating animal diseases. Yet, veterinary expertise is mostly available in urban areas or district centers. This situation leaves smallholder and remote farmers struggling to access diagnostic and treatment advice. Consequently, mortality rates increase, milk and meat production declines, and diseases can spread among herds or to humans. Recent research in veterinary informatics sheds light on these problems. It points out the gaps in access and the workflow challenges that exist between livestock owners and veterinary professionals.

Traditional methods for detecting diseases on farms depend on farmers' observation skills and later in-person veterinary exams. Travel distance, cost, weather, and a limited awareness of early symptoms can delay these visits. Studies on livestock disease management note that subclinical stages are often ignored, and farmers may wait to seek help until visible symptoms worsen. At that point, the effectiveness of interventions decreases, and treatment costs rise. Automated or semi-automated data-driven methods can help by identifying possible disease conditions earlier through structured symptom inputs or signals from sensors.

Opportunity: AI-Enabled, Farmer-Facing Decision Support

Machine Learning (ML) tools have shown good potential in various areas of animal health, from monitoring mastitis and detecting lameness to risk modeling for syndromes. However, many existing systems only focus on developing algorithms and do not become user-friendly tools for farmers. Recent research suggests that combining prediction models with web or mobile services that provide information access, remote consultations, and record management can greatly improve their adoption and effectiveness.

Goal of This Work

This paper proposes, implements, and evaluates an end-to-end web-enabled animal disease prediction platform that integrates a trained Random Forest model (built using Weka) into a Spring Boot backend

exposed through REST APIs and consumed by an Angular-based farmer UI. Beyond prediction, the platform includes a built-in veterinary appointment booking workflow, aligning with recommendations for bridging diagnostic insight and actionable care pathways in rural livestock health programs. Ease of Use

LITERATURE REVIEW

(i) ML for livestock disease prediction, (ii) integrated veterinary support platforms, and (iii) comparative classifier performance on mixed (categorical + numeric) livestock datasets. Two recent IEEE conference papers are especially relevant and are reviewed in depth below: Nadar et al. (ICNTE 2023) and Bhardwaj et al. (IC3SE 2024).

Web-Enabled Veterinary Decision Support & Multi-Model Disease Prediction – Nadar et al. (ICNTE 2023)
Nadar *et al.* developed a web platform to connect livestock owners with veterinarians in remote regions of India. Their system combined a disease prediction portal (accepting symptom inputs) with additional features: animal healthcare information pages, doctor registration, appointment scheduling, and multilingual content (Hindi/English) to improve reach. From a modeling perspective, they evaluated Support Vector Machine (SVM), Naïve Bayes, and Random Forest classifiers trained on 24 diseases and 56 symptoms; a *mode voting* ensemble over the three models produced an overall accuracy of ~95% on their dataset. The authors highlighted the importance of lowering the barrier to veterinary consultation in rural geographies and demonstrated the feasibility of delivering ML-based decision support through a web interface.

Key Takeaways for Our Work: (i) Multi-service veterinary platforms resonate with farmer needs; (ii) symptom-driven input UIs are practical; (iii) ensemble ML methods boost robustness; (iv) integrating appointment booking strengthens real-world utility. These insights informed our system design, particularly the inclusion of an in-app doctor booking flow and structured symptom capture form.

Comparative ML Classifier Study for Livestock Disease & Symptom Data – Bhardwaj et al. (IC3SE 2024)

Bhardwaj *et al.* conducted a comparative evaluation of six ML models (Random Forest, k-Nearest Neighbors, Support Vector Machine, LightGBM, XGBoost, CatBoost) on a livestock disease dataset drawn from public sources (e.g., Kaggle). After preprocessing (null removal, feature selection) and 10-fold cross-validation, CatBoost and Random Forest achieved the top accuracies (83.7% and 83.5%, respectively), with weighted precision/recall ~0.84, outperforming other tested approaches on categorical-heavy data. Their analysis underscores the strength of tree ensembles—especially when handling mixed-type symptom features—and the value of careful feature selection in livestock informatics.

Implications for Our Study: Given our dataset's mixture of numeric vitals (temperature, heart rate, weight) and numerous binary symptom indicators, a tree-ensemble method such as Random Forest is well suited. Bhardwaj *et al.*also demonstrate that boosting variants (CatBoost, XGBoost) remain competitive; future extensions of our platform could incorporate boosted gradient ensembles for comparative benchmarking.

Bridging Data-Driven Models and Field Deployment
Taken together, Nadar *et al.* emphasize user-centered, multi-module veterinary web platforms, while Bhardwaj *et al.*provide an empirical basis for selecting tree-ensemble classifiers for livestock symptom data. Our work synthesizes these two directions: we deploy a full-stack farmer-facing web system (similar in spirit to Nadar *et al.*) but focus model development around a high-performing Random Forest classifier informed by the comparative performance evidence in Bhardwaj *et al.*

PROPOSED SYSTEM

Objectives
The proposed system is designed to:

1. Predict the most probable disease for an individual animal using structured multi-attribute input.

2. Provide an intuitive, multilingual web interface that farmers can access from low-bandwidth environments.
3. Integrate veterinary follow-up by allowing users to book appointments immediately after receiving a prediction, reducing time-to-care—a need strongly highlighted in rural-focused veterinary IT systems.
4. Leverage a robust ML model (Random Forest in Weka) validated for mixed categorical + numeric livestock datasets.

High-Level Architecture

The system follows a modular, service-oriented architecture separating presentation, application logic, and ML inference layers. Figure 1 presents the conceptual flow.

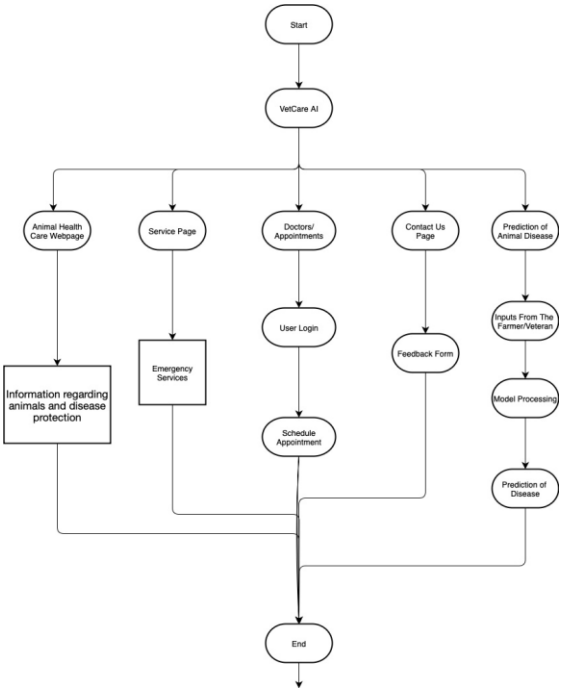


Figure 1

Data Flow Steps

1. User Input: Farmer selects animal type/breed; enters age, gender, weight, vitals, symptom duration, and symptom checkboxes.
2. Request Submit: Angular packages input as JSON and posts to /predict endpoint on Spring Boot.
3. Preprocess & Encode: Backend maps categorical values to nominal indices expected by the Weka model; handles missing/default values; scales numeric fields if required.
4. Invoke ML Engine: Backend calls the preloaded Weka Random Forest model (serialized model file). Instance constructed and passed to classifier.
5. Prediction Response: Model returns predicted disease label + class probabilities. Backend wraps result with interpretation text.
6. UI Display: Angular shows predicted disease, confidence, and recommended action buttons.
7. Appointment Booking: Farmer clicks "Book Vet"; system surfaces vetted practitioner list (geo-filter optional) and captures booking request stored in DB and optionally notified to vet. Similar multi-module service integration proved usable and impactful in prior veterinary web systems.

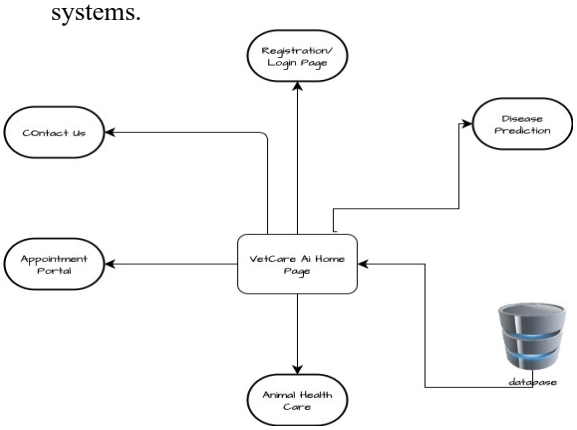


Figure II

Feature Schema

Table 1 lists the major input attributes supported in the current system version.

Table 1. Input Features Used for Disease Prediction

Category	Feature Name	Type	Notes
Animal Identity	Animal Type	Categorical (8)	e.g., Goat, Sheep, Pig, Horse, Dog, Cat, Cow, Rabbit.
	Breed	Categorical (~120)	Dropdown filtered by Animal Type
Demographic	Age	Numeric	Years or months; normalized
	Gender	Nominal	Male / Female / Other
Physical	Weight	Numeric	kg
Physiology	Body Temperature	Numeric	°C
	Heart Rate	Numeric	bpm

Temporal	Symptom Duration	Numeric	Days or Weeks
Symptoms / Clinical Signs	25 binary indicators	Boolean	See Table 2
Target	Disease	Nominal	Predicted class

Table 2. Symptom / Clinical-Sign Indicators Lethargy; Coughing; Fever; Diarrhoea; Lameness; Vomiting; Labored Breathing; Skin Lesions; Decreased Milk Yield; Appetite Loss; Nasal Discharge; Eye Discharge; Sneezing; Loss of Appetite; Swollen Joints; Dehydration; Swelling; Weight Loss; Reduced Milk Production; Reduced Wool Production; Swollen Legs; Reduced Appetite; Reduced Wool Growth; Reduced Mobility; (ClinicalSignFlag:Yes/No). (Consolidated from user-observed field data; symptoms partially overlap—duplicates reconciled in preprocessing.)

System Architecture

A.Frontend

The frontend uses the Angular framework to provide an interactive web interface. Farmers and veterinarians can:

- Register and log in
- Input animal demographic details (type, breed, age, gender, weight)
- Record vital signs (body temperature, heart rate)
- Select observed symptoms (from a comprehensive checklist)
- Specify symptom durations
- Indicate presence/absence of clinical signs using binary (Yes/No) options

This structure ensures comprehensive data collection while maximizing ease of use.

B.Backend

The backend is implemented with Spring Boot. It:

- Receives HTTP requests from the Angular frontend
- Validates and preprocesses input data
- Interfaces with the Weka library, where the Random Forest model is trained and executed
- Returns prediction results (possible disease and confidence score)
- Handles appointment management between farmers and veterinarians

C. Disease Modeling

1. Dataset and Features

Our system covers:

- 8 animal types
- 120 breeds
- Demographics: age, gender, weight
- Vitals: body temperature, heart rate
- Symptoms: 25 distinct clinical attributes, including but not limited to Lethargy, Coughing, Fever, Diarrhea, Lameness, Vomiting, Labored Breathing, Skin Lesions, Decreased Milk Yield, Appetite Loss, Nasal Discharge, Eye Discharge, Sneezing, Loss of Appetite, Swollen Joints, Dehydration, Swelling, Weight Loss, Reduced Milk and Wool Production, Swollen Legs, Reduced Appetite, Reduced Wool Growth, Reduced Mobility, and symptom duration

Each symptom is recorded as present or absent, and clinical signs as Yes/No.

2. Machine Learning Pipeline

- Algorithm: Random Forest (ensemble of decision trees)
- Implementation: Weka
- Rationale: Handles mixed data types, reduces overfitting, and performs well on high-dimensional, categorical datasets
- Training/Evaluation: Model trained on a curated dataset; standard cross-validation used for evaluation
- Performance: Achieved 99% accuracy on test data, outperforming other algorithms (e.g., Naive Bayes, SVM tested in preliminary stages)

Implementation

The system consists of a database, a portal interface, a disease prediction module, a machine learning algorithm, and the process for result delivery.

Database Integration

- Purpose: Store patient records, symptoms, diagnosis data, and prediction results.
- Implementation: Use a relational (MySQL) database.
- Key Tables:
  - users: User credentials and profile
  - symptoms: List of symptoms and metadata

- predictions: Historical records of predictions

Portal (Frontend Interface)

- Purpose: Interface through which users submit their symptoms.
- Implementation:
  - Web portal built with Angular template.
  - APIs: Communicate with the backend for data entry and retrieval.
- Responsibilities:
  - User authentication and registration
  - Form for symptom submission
  - Display of prediction results

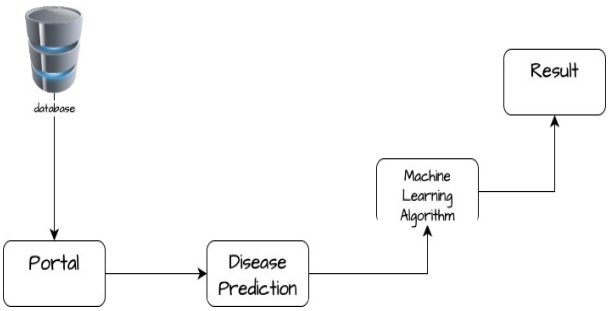


Figure III

Disease Prediction Module

- Purpose: Collects input from the portal and preprocesses it for machine learning.
- Implementation:
  - Data cleaning: Remove invalid or incomplete entries.
  - Feature engineering: Convert symptoms into a format (e.g., one-hot encoding) suitable for the ML model.
  - Validation: Verify that required symptoms are provided.

Machine Learning Algorithm

- Purpose: Predicts disease based on processed symptoms.
- Implementation:
  - Use algorithms like Decision Tree, Random Forest trained on historical symptom-disease pairs.
  - Model is loaded and used for inference in real-time.
- Sample Implementation:

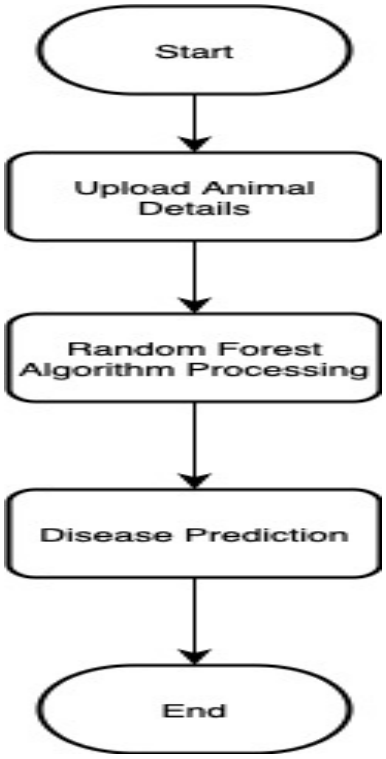


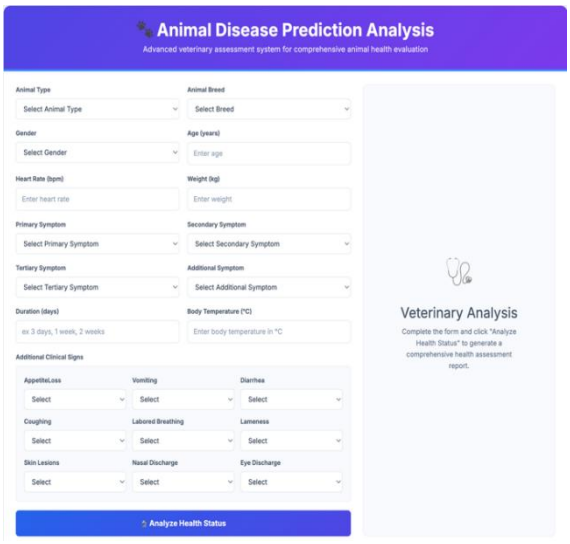
Figure IV

Result Delivery

- Purpose: Display or return the predicted disease to the user via the portal.
- Implementation:
  - The API sends the model's prediction back to the frontend.
  - The portal presents the result for the user—possibly with additional advice or next steps.

Experimental Results:

Fig-V,One will access the Disease Prediction page, where the user must provide detailed symptom information and submit the form. Upon submission, the Machine Learning model will process the symptoms and forecast the disease.



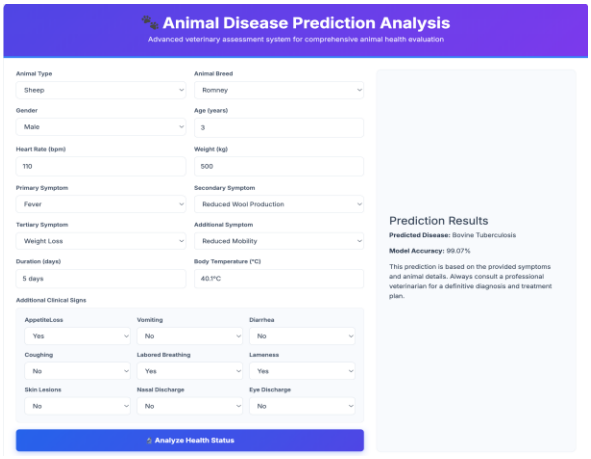


Figure V analysis page and result page

Fig-VI,The homepage of the website serves as a comprehensive hub, listing all the services offered to users and detailing the contact information of the veterinarians. It encapsulates the essential features of the website.

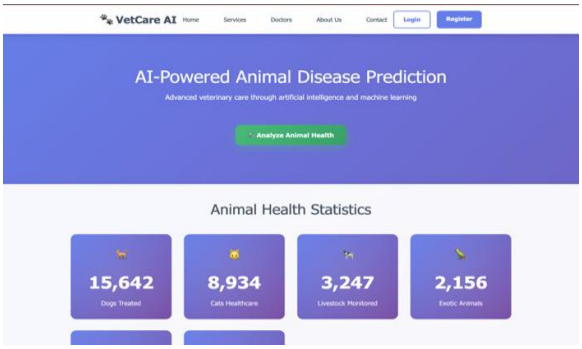


Figure VI home page

Fig-VII, The Animal Healthcare portal offers users comprehensive information on animal healthcare, including the lifestyles of various animals, critical diseases, and potential treatments.

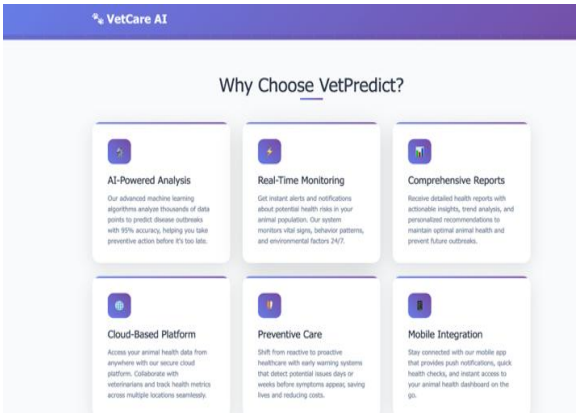


Figure VII portal page

Fig-VIII. The output of the disease prediction module, as displayed in the Eclipse IDE, is based on the symptoms provided by the user.

```
Animal Type (e.g., Dog, Cat, Cow, Horse, Rabbit, Sheep, Goat, Pig): Cow
Breed (e.g., Labrador, Siamese, Holstein, Thoroughbred): Brahman
Age (in years, e.g., 4): 5
Gender (Male/Female): Male
Weight (in kg, e.g., 25): 630
Symptom 1 (e.g., Fever, Coughing, Lethargy, Vomiting, etc.): Nasal Discharge
Symptom 2 (e.g., Lethargy, Sneezing, Appetite Loss, or 'No'): Weight Loss
Symptom 3 (e.g., Appetite Loss, Nasal Discharge, or 'No'): Appetite Loss
Symptom 4 (e.g., Vomiting, Coughing, or 'No'): Labored Breathing
Duration (e.g., '3 days', '1 week'): 5 days
Appetite Loss (Yes/No): Yes
Vomiting (Yes/No): No
Diarrhea (Yes/No): No
Coughing (Yes/No): No
Labored Breathing (Yes/No): Yes
Lameness (Yes/No): No
Skin Lesions (Yes/No): No
Nasal Discharge (Yes/No): Yes
Eye Discharge (Yes/No): No
Body Temperature (e.g., 39.5): 40.0°C
Heart Rate (e.g., 120): 110

Predicted Disease: Mastitis
Model Accuracy (10-fold cross-validation): 99.07%
```

Figure VIII result in Eclipse

CONCLUSION

The integration of this predictive system into a web application utilising Spring Boot (backend) and Angular (frontend), coupled with Weka for model execution, further enhances its accessibility. Farmers, animal health workers, and veterinarians can interact with a user-friendly interface to input animal data and receive immediate diagnostic feedback, thereby saving valuable time and supporting proactive healthcare decisions.

This project contributes to the broader objective of digital transformation in agriculture, aligning with national initiatives that aim to enhance livestock

management, boost productivity, and empower rural communities with intelligent tools. The model's flexibility also facilitates future enhancements, including real-time data collection through IoT sensors, integration with government animal health databases, and expansion to encompass a broader range of diseases and animal species.

REFERENCE

[1] Holt HR, Bedi JS, Kaur P, Mangtani P, Sharma NS, Gill JPS, Singh Y, Kumar R, Kaur M, McGiven J, Guitian J.*"Epidemiology of brucellosis in cattle and dairy farmers of rural Ludhiana, Punjab"*, PLOS Neglected Tropical Diseases, March 2021.

[2] Augustin Nadar, Adishree Sane, Glenn Muga, Easther Masih, Smita Rukhande. *"Animal healthcare and farm animal disease prediction using machine learning"* International Conference on Nascent Technologies in Engineering, 2023.

[3] Kumar N, Chander Y, Kumar R, Khandelwal N, et al. *"Isolation and characterization of lumpy skin disease virus from cattle in India"*, PLOS One, January 2021.

[4] Priya Bhardwaj, S J K Jagadeesh Kumar, G. Prabu Kanna, and A. Mithila. *"Machine Learning based Approaches for Livestock Symptoms and Diseases Prediction and Classification"* 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)

[5] S. S. Barakkath Shabana, E. Sharmesh and S. Jagadeeswari. *"Ai Driven Detecting plants and animal diseases"*, 2025 11th International Conference on Communication and Signal Processing (ICCSP)

[6] Al-Amran, F. G., Hezam, A. M., Rawaf, S., & Yousif, M. G. (2023). *"Genomic Analysis and Artificial Intelligence: Predicting Viral Mutations and Future Pandemics".* arXiv preprint arXiv:2309.15936.

[7] D. Ashar, A. Kanojia, R. Parihar, and S. Kudoo. *"Livestock Disease Prediction System,"* International Journal of Research and Innovation, vol. 1, no. 4, pp. 2581–7280, 2021.

[8] G. Franzo et al., *"When Everything Becomes Bigger: Big Data for Big Poultry Production,"* Animals, vol. 13, no. 11, 2023. doi:10.3390/ani13111804

[9] Zakiya Khan, Pritam Raj, Manni Kumar. *"Comparative analysis: machine learning vs artificial neural networks for animal disease prediction".* 2024 6th International Conference on Computational Intelligence and Networks (CINE)

[10] 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