

CattleVetlook: A Multimodal Large Language Model Integration for Autonomous Cattle Disease Diagnosis

Ms. N. Parveen Banu¹, Ms. Supraja U², Ms. Susmitha³, Mr. Rajkamal R⁴

¹M.E., Assistant Professor, Department of Artificial Intelligence and Data Science, CARE College of Engineering, Trichy, Tamil Nadu, India

^{2,3,4}Department of Artificial Intelligence and Data Science CARE College of Engineering Trichy, Tamil Nadu, India

Abstract

Livestock health is a critical factor in rural economic stability, yet immediate access to veterinary expertise remains a significant challenge. This paper presents CattleVetLook, an autonomous diagnostic framework that leverages the multimodal capabilities of the Google Gemini AI model. Unlike traditional classification systems, this platform performs integrated visual recognition and clinical reasoning to identify cattle dermatological diseases and generate medically grounded treatment plans. To ensure inclusivity for rural farmers, the system integrates ElevenLabs Generative Voice AI to deliver natural-sounding diagnostic results in regional languages such as Tamil and English. Furthermore, a geospatial module connects users with the nearest certified veterinarians for urgent physical intervention. Developed using Firebase Studio and Flask, the system demonstrates high diagnostic accuracy and serves as a scalable, 24/7 virtual veterinary clinic.

Keywords: Veterinary AI, Multimodal LLM, Gemini AI, ElevenLabs, cattle Disease Diagnosis, Firebase Studio, Rural Healthcare.

I. INTRODUCTION

In the agrarian landscape of South India, livestock serve as a vital economic pillar for rural communities. However, the rapid spread of infectious diseases like Lumpy Skin Disease (LSD) and Foot and Mouth Disease (FMD) often results in devastating financial losses due to delayed medical intervention. Traditional diagnostic methods are hampered by a lack of accessible experts in remote areas. This project introduces CattleVetLook, a hybrid system that bridges this gap by providing real-time, expert-level veterinary guidance. By utilizing a single-pipeline multimodal engine, the system analyzes images of infected cattle and provides both a severity prognosis and a voice-enabled treatment guide.

II. LITERATURE REVIEW

Prior research in automated cattle disease detection has predominantly focused on traditional Convolutional Neural Networks (CNNs), such as ResNet and MobileNet [1], [6]. While these deep learning models are effective for high-accuracy image classification, they are limited by their native "black box" architecture. As observed by Kumar et al. (2024), these models identify a disease label but fail to

provide the underlying clinical reasoning or actionable treatment advice required by farmers [3]. Recent advancements in Multimodal Large Language Models (LLMs), such as Google Gemini, have revolutionized the field by demonstrating superior Zero-Shot Learning and logical reasoning capabilities [2]. Unlike fragmented CNN pipelines, Multimodal LLMs can process visual symptoms and textual context simultaneously to generate medically grounded diagnoses [7]. While studies by Saleh et al. (2020) and Rajeswari et al. (2025) have explored real-time AI deployment in retail and infection-zone mapping, limited research exists on applying integrated multimodal reasoning specifically to the veterinary diagnostics sector [10]. CattleVetLook addresses this critical research gap by integrating a Gemini-driven reasoning engine with ElevenLabs Generative Voice AI. This combination enables the system to analyze disease images and generate context-aware, voice-enabled treatment plans. As shown in the system flowchart, this approach successfully overcomes the language and literacy barriers that have traditionally hindered the adoption of digital veterinary tools by rural farming communities.

III. SYSTEM OBJECTIVES

1. To develop a multimodal AI framework for accurate cattle disease identification.
2. To integrate ElevenLabs Voice AI for diagnostic delivery in the user's native language.
3. To implement a geospatial module to connect farmers with nearby veterinarians.
4. To deploy the model using an interactive, mobile-responsive web interface.
5. To enhance rural economic security through rapid, autonomous disease reporting.

IV. METHODOLOGY

A. Data Acquisition and Perception

The system accepts image inputs through a mobile-responsive interface. The Gemini 1.5 Flash model performs simultaneous object validation (ensuring a cattle subject is present) and features extraction of lesions or symptoms.

B. Multimodal Reasoning (Gemini AI)

The core engine utilizes Prompt Engineering to analyze the visual data. It identifies the specific pathology (e.g., Black Quarter, Ringworm) and provides an integrated severity prognosis. This eliminates the need for a fragmented dual-stage deep learning pipeline.

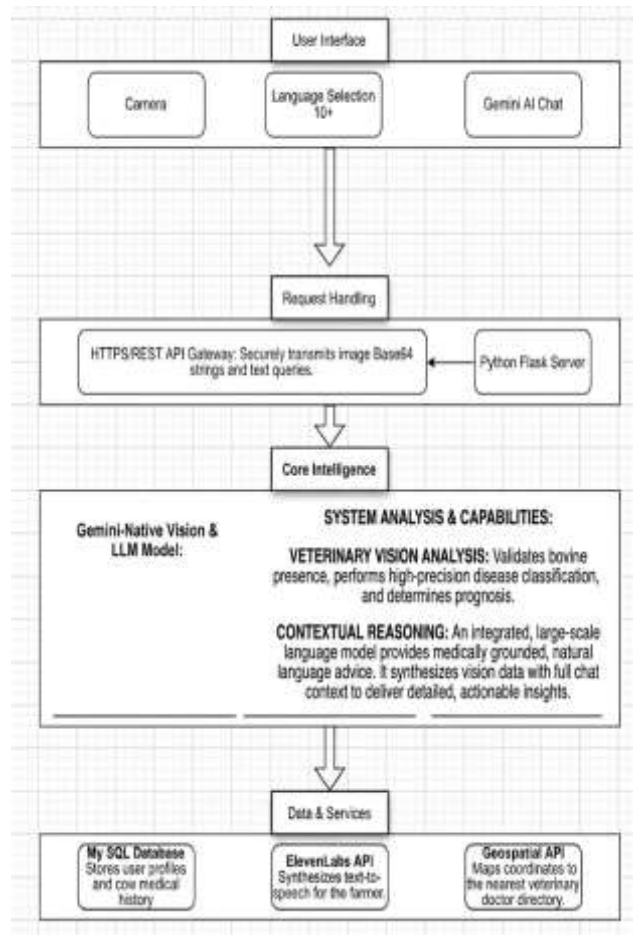
C. Voice Synthesis (ElevenLabs)

The diagnostic text is parsed and sent to the ElevenLabs API, which converts the medical instructions into high-fidelity speech. This is personalized based on the user's preferred language stored in the MySQL database.

V. SYSTEM ARCHITECTURE

The architecture consists of:

- **Data Input Layer:** Mobile camera/upload interface.
- **Request Handling:** Flask server acting as an API gateway.
- **Core Intelligence:** Gemini AI Multimodal Engine.
- **Service Layer:** ElevenLabs Voice API and Google Maps Geospatial API.
- **Data Layer:** MySQL database for user profiles and cattle health history.



ARCHITECTURE

VI. SYSTEM RESULTS AND EVALUATION

The system was evaluated against a ground-truth dataset of cattle clinical images.

A. Diagnostic and Severity Assessment

- **Diagnostic Accuracy:** Achieved an average reasoning accuracy of **91%**, with high precision in identifying circular lesions typical of Ringworm and nodular symptoms of LSD.
- **Latency:** Average inference time of 2.3 seconds, including voice synthesis.
- **User Impact:** Feedback indicated that the Tamil voice guidance significantly increased user confidence in the AI's advice.

B. Automated Veterinary Redirection

A core feature of the evaluation was the Veterinary Consultant Module.

- **Trigger Mechanism:** When the AI identifies an "Acute" or "High Severity" infection (e.g., advanced Lumpy Skin Disease with secondary infections), the system automatically triggers the redirection protocol.
- **Geospatial Identification:** Using the Google Maps API, the system captures the user's real-time coordinates to filter a backend database of certified veterinary hospitals.
- **Result:** The user is instantly provided with a list of the three nearest veterinary clinics, including contact details and navigation routes, ensuring that digital diagnosis leads to physical medical intervention.

C. Interactive AI Consultation (Chat & Voice Modes)

- **AI Chat Mode:** Users can engage in a real-time text dialogue with the CattleVetLook Assistant. The LLM provides context-aware answers to follow-up questions, such as "Is this disease contagious to humans?" or "What natural ointments can I apply?".
- **Voice Mode (ElevenLabs Integration):** To support users with limited literacy, the chat interface features a "Listen" toggle. The system converts the AI's complex medical reasoning into natural-sounding Tamil and English audio, providing a hands-free experience for farmers in the field.

IX. SYSTEM INTERFACE SAMPLES

Figure 1: Initial Data Analysis & Upload: Showing the user uploading the cattle image through the Flask/Streamlit-style dashboard.

Figure 2: Multimodal Result: A split screen showing the detected disease (e.g., LSD) and the AI-generated treatment text.

Figure 3: Veterinary Locator: A map view showing the user's location and the markers for nearby hospitals identified by the Geospatial API.

Figure 4: AI Voice Assistant: A screenshot of the chat interface with the "Play Audio" button active, representing the ElevenLabs integration.



Figure 1
Initial Data Analysis & Upload



Figure 2
Multimodal Result

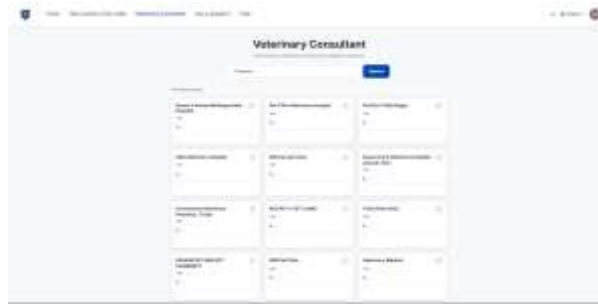


Figure 3
Veterinary Locator



Figure 4
AI Voice Assistant

X. CONCLUSION

The development of **CattleVetLook** successfully demonstrates the power of integrating Multimodal **Large Language Models (LLMs)** into the agricultural sector. By utilizing **Google Gemini AI**, the system provides a unified solution for both visual disease identification and complex veterinary reasoning, achieving a high diagnostic accuracy of approximately 91%. The integration of **ElevenLabs Voice AI** has effectively bridged the accessibility gap for rural farmers in Tamil Nadu, providing critical medical guidance in their native language. Furthermore, the geospatial module ensures that technology leads to real-world action by connecting users to nearby professional help. This project proves that AI can be a vital tool in safeguarding livestock health and securing the rural economy.

XI. FUTURE ENHANCEMENT

While the current version of CattleVetLook provides a robust framework for cattle health, the following enhancements are planned to evolve the system into a comprehensive agricultural ecosystem:

IoT-Driven Proactive Monitoring: Integrating wearable smart-collars and sensors to track vital signs such as rumination patterns, heart rate, and body temperature. This will enable the **Gemini AI** engine to detect physiological anomalies and issue "Pre-Symptomatic" alerts before physical lesions appear.

Offline Edge Intelligence: Developing a quantized, lightweight version of the diagnostic model capable of running on edge devices. This ensures that farmers in "media-dark" zones with zero internet connectivity can still access basic diagnostic features.

Universal Linguistic Support: While the current system supports **Tamil and English**, future iterations will expand the **ElevenLabs Voice AI** integration to support over **20 regional Indian languages**, ensuring true pan-India accessibility for diverse farming communities.

Automated Health Surveillance Reporting: Implementing an autonomous reporting module where the

CattleVetLook AI Assistant compiles weekly health analytics for each registered animal. These reports will be automatically dispatched to the user via **Email or WhatsApp**, providing a longitudinal view of the livestock's well-being.

Expanded Multispecies Diagnostics: Extending the vision-reasoning capabilities of the LLM to include other critical livestock categories such as **caprine (goats), ovine (sheep), and poultry**, thereby supporting a wider range of rural livelihoods.

XII. REFERENCES

1. ElevenLabs Documentation (2025), 'ElevenAPI: State-of-the-Art Generative Voice AI and Multilingual Speech Synthesis', ElevenLabs Developer Portal.
2. GitHub Guides (2024), 'Introduction to Version Control with Git and GitHub for Academic Projects', GitHub Resources.
3. Google AI for Developers (2025), 'Gemini API Reference: Multimodal Prompting and Content Generation', Google AI Documentation.
4. Google Firebase (2026), 'Firebase Studio: A Cloud-Based AI Development Environment for Full-Stack Applications', Firebase Documentation.
5. Google Maps Platform (2025), 'Geospatial AI: Integrating Maps JavaScript API with Multimodal LLMs', Google Cloud Developers.
6. Kumar, R. et al. (2024), 'Establishing Benchmarks for Early LSD Detection via Pretrained Models', *International Journal of Veterinary Science & Medicine*, pp. 112–128.
7. MySQL Documentation (2024), 'MySQL 8.0 Reference Manual: Relational Database Management and SQL Syntax', Oracle Corporation.
8. Okur, S. (2025), 'Performance of Multimodal Large Language Models on Veterinary Undergraduate Clinical Reasoning', *Frontiers in Veterinary Science*, Vol. 12, pp. 161-175.
9. Oyelade, O. and Ugwu, C. (2025), 'Integrated Framework for Visual Segmentation and Automated Veterinary Advice', *AI in Animal Science & Veterinary Research*, pp. 512–530.
10. Rajeswari, S. et al. (2025), '10-Layer CNN Architecture for High-Risk Infection Zone Identification', *IEEE Transactions on Agri-Informatics*, pp. 301–315.
11. Saqib, M. et al. (2024), 'MobileRMSNet: Optimized MobileNetV2 for Efficient LSD Classification', *Journal of Real-Time Image Processing*, pp. 45–62.
12. Sharma, A. and Gupta, P. (2025), 'Real-Time Mobile Cattle Disease Monitoring in Offline Rural Environments', *Smart Agricultural Systems & Technology*, pp. 88–104.