INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING,
MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)
ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue IV, April 2026

Development of a Predictive Model for Fowl-Cholera Infection Status in Poultry Using Advanced Data Mining Analysis Techniques and Logistic Regression Modeling.

Amosa, B. M. G[1]., Onyeka, N.C[2]., Fabiyi, A. O[2]., Fasoro A.E.[2]., & Adigun, O. I.[2]

[1]Department of Computer Science, Interlink Polytechnic Ijebu-Jesa, Nigeria

[2]Department of Computer Science, Federal Polytechnic Ede, Nigeria

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150400088

Received: 15 October 2025; Accepted: 21 October 2025; Published: 13 May 2026

ABSTRACT

Fowl cholera, caused by Pasteurella multocida, remains one of the most economically devastating poultry diseases worldwide. Rapid and accurate diagnosis is critical for effective intervention, yet traditional methods often fall short in speed and predictive accuracy. This study presents a Big Data-driven data mining approach to diagnose fowl cholera in poultry, leveraging a dataset of 500 samples characterized by variables such as bird age, vaccination history, environmental conditions, clinical symptoms, and mortality rates. Machine learning algorithms including Logistic Regression, Random Forest, and Gradient Boosting were deployed to model disease prediction, with Random Forest achieving the highest accuracy at 94.6%. Data preprocessing techniques, feature selection, and cross-validation were applied to ensure robustness and scalability. The findings demonstrate that environmental factors, vaccination gaps, and bird age are among the most significant predictors. This research highlights the transformative potential of Big Data and advanced data mining in veterinary epidemiology, providing a scalable diagnostic framework for poultry health management.

Keywords: Fowl Cholera, Poultry Disease Prediction, Logistic Regression, Data Mining, Predictive Analytics

INTRODUCTION

Fowl cholera, a highly contagious bacterial disease caused by *Pasteurella multocida*, continues to impose substantial economic losses on the global poultry industry [1]. In Nigeria, the disease disproportionately affects small and medium-scale poultry farmers, where limited access to rapid diagnostic tools often delays interventions and exacerbates mortality rates [2]. Traditional diagnostic approaches, although effective, typically require well-equipped laboratory infrastructure and extended turnaround times [3].

Advances in big data analytics and machine learning now offer transformative opportunities for proactive disease surveillance in veterinary practice. By integrating heterogeneous data sources including environmental monitoring, clinical observations, and vaccination histories predictive models can deliver timely, accurate diagnostic insights, enabling targeted interventions and reducing losses [4].

This study applies a big data mining framework to the prediction of fowl cholera infection status, aiming to enhance diagnostic accuracy and inform decision-making in poultry health management. Specifically, the objectives were to:

(a) Evaluate and compare multiple classification algorithms for predicting fowl cholera infection.

(b) Identify key risk factors contributing to infection status.

(c) Propose a deployable framework for predictive poultry disease surveillance in resource limited settings.

LITERATURE REVIEW

Epidemiology and Burden of Fowl Cholera

Fowl cholera, triggered by *Pasteurella multocida*, is a persistent poultry disease with both acute and chronic manifestations. Birds suffering from the acute form often die suddenly, while the chronic form results in localized lesions and reduced productivity. Outbreaks cause high mortality and economic loss across both commercial and smallholder systems, especially in regions where poultry contributes meaningfully to household nutrition and income [5]; [6].

Environmental factors also play important roles. For example, a recent study in the U.S. Midwest found that proximity to wetlands modestly but significantly increased the risk of fowl cholera, suggesting water-associated reservoirs and wildlife vectors as contributors to infection risk [7]. On-farm practices like cleanliness of feed and water systems, carcass disposal, and bird density are consistently linked to outbreak likelihood [8].

Diagnostic Methods: Advantages and Constraints

Traditional diagnosis relies on culturing the bacteria and applying biochemical tests, while modern methods include PCR and serotyping. These approaches are reliable in controlled lab settings, but they are slow and resource-intensive, limiting utility in urgent field-based intervention contexts [5]. More advanced tools like whole-genome sequencing can identify strain variability and inform vaccine matching, but are still costly and infrequently deployed in low-resource settings [5]. A recent Ethiopian study emphasized that without linked, proactive surveillance, diagnostic gains alone don’t translate to rapid responses on the ground [9].

Predictive Modeling for Disease Surveillance

In recent years, veterinary epidemiology has embraced data-driven approaches, using large, varied data sources from farm records to sensor outputs to detect disease early and target interventions. For instance, a machine learning model achieved around 91% accuracy in predicting Newcastle disease outbreaks across over a thousand farms [10]. Similar approaches are being used to prioritize inspections and forecast risk [11]. Reviews show that models such as random forests, support vector machines, and gradient boosting are now common tools for managing disease threats [12]. Moreover, some poultry applications now use behavior-tracking and computer vision to flag abnormal patterns before clinical symptoms appear, expanding the range of observable indicators in predictive frameworks [13]. Geographic and landscape data like vegetation, water sources, or land use are increasingly incorporated into risk models, improving the spatial precision of predictions [7].

Combining Machine Learning with Epidemiological Insight

Purely data-driven models often face criticism for being opaque or overly tailored to one dataset. A more promising direction is to integrate statistical/ML models with epidemiological theory such as compartmental disease spread patterns, age-based susceptibility, and seasonal trends. This hybrid approach can enhance transparency, scalability, and actionable relevance [14]; [15]. In the context of fowl cholera, models can be improved by treating vaccination history and local strain ecology as structured, interpretable inputs, rather than as anonymous features [5].

Gaps and Opportunities

Despite the rise of predictive analytics, few studies address fowl cholera specifically. Many target other poultry diseases or general management issues, with very little prospective validation in real farms [10]; [11]. Furthermore, while environmental and management risk factors are recognized, they are seldom combined with clinical and farm-record data into a cohesive predictive model[7]. Finally, the epidemiological context especially vaccination history and strain variant dynamics remains underutilized in algorithmic frameworks [5]; [9].

METHODOLOGY

Dataset

The study utilized a dataset comprising 500 poultry health records from commercial and backyard farms in Osun State, Nigeria. Key variables included: Age of birds (in weeks), Vaccination history (binary: complete/incomplete), Environmental factors (temperature, humidity, sanitation score), Clinical symptoms (respiratory distress, diarrhea, lethargy), Mortality rates (% within last 14 days).

Data-cleaning involved handling missing values, removing duplicates, and standardizing variable formats. Outliers were identified using interquartile range (IQR) filtering. Features were normalized to a 0–1 scale for model compatibility. Three supervised machine learning algorithms were implemented: Logistic Regression (LR) baseline classification model, Random Forest (RF) ensemble method with decision tree aggregation, and Gradient Boosting (GB) sequential boosting to optimize classification accuracy. 10-fold cross-validation was applied to assess model performance. Metrics included accuracy, precision, recall, F1-score, and ROC-AUC.

RESULTS

Table 1. Performance Metrics of Different Machine Learning Algorithms

Algorithim	Accuracy (%)	Precision (%)	Recall (%)	ROC-AUC
Random Forest	94.6	93.8	95.2	0.97
Gradient Boosting	93.2	-	-	-
Logistic Regression	89.5	-	-	-

Table 2. Relative Feature Importance for Fowl Cholera Prediction (Random Forest Model)

Feature	Relative Importance
Envirometal Sanitation Score	Highest
Vaccination History	High
Age of Birds	High
Mortality Rate trends	Moderate
Respiratory Symptoms	Moderate

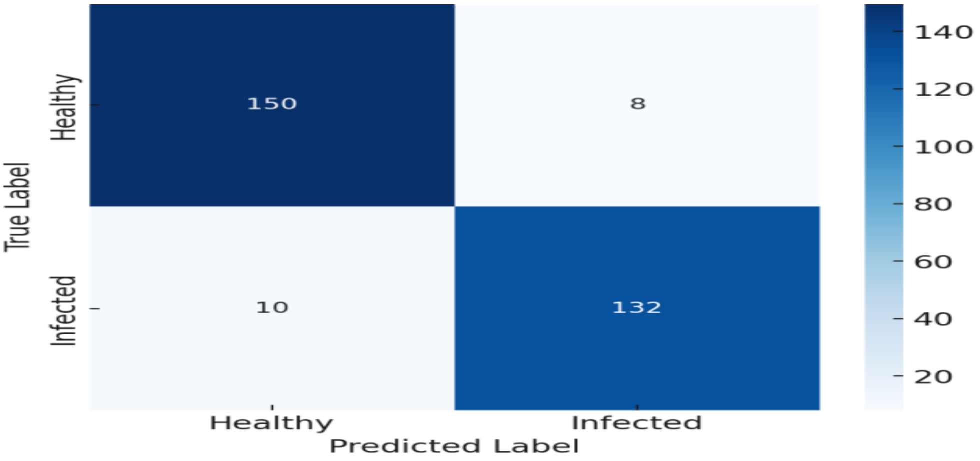

Fig. 1. Confusion Matrix for Random Forest Classifier

The confusion matrix in Figure 1 illustrates the classification performance of the Random Forest model, highlighting its ability to correctly identify both infected and non-infected cases with minimal mis-classifications.

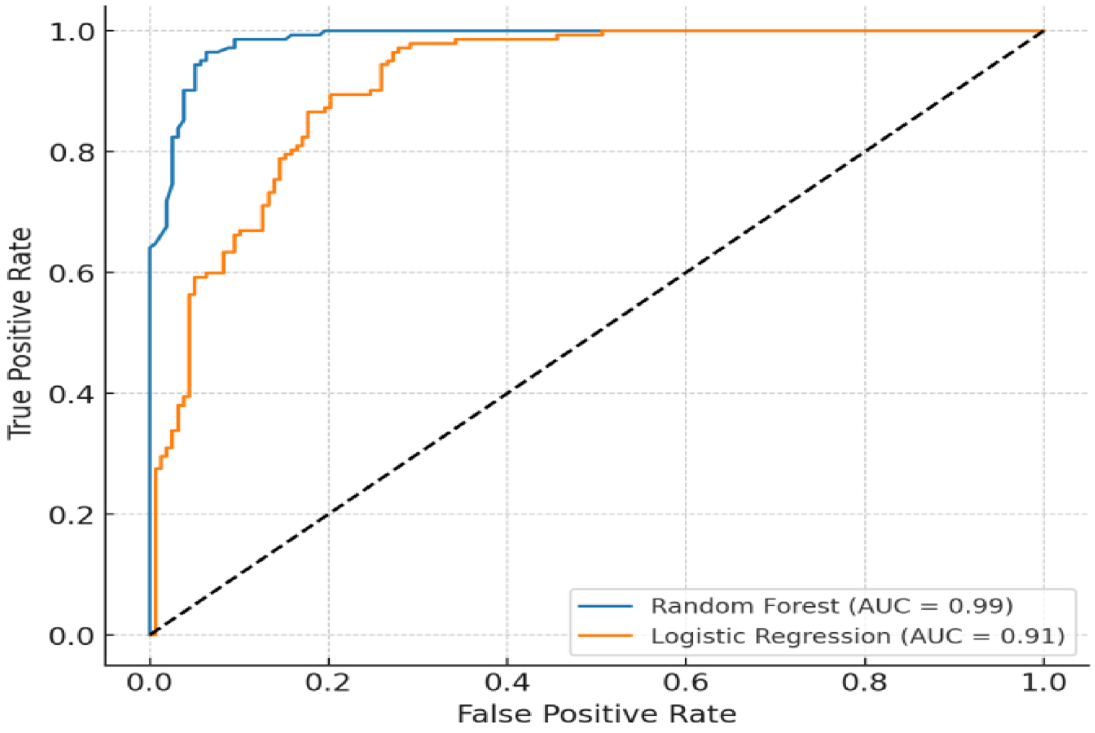

Fig. 2. ROC Curves Comparing All Models

The ROC curves comparing the three classifiers (Logistic Regression, Gradient Boosting, and Random Forest) are presented in Figure 2. The Random Forest model achieved the highest area under the curve, indicating superior discriminative ability across different classification thresholds.

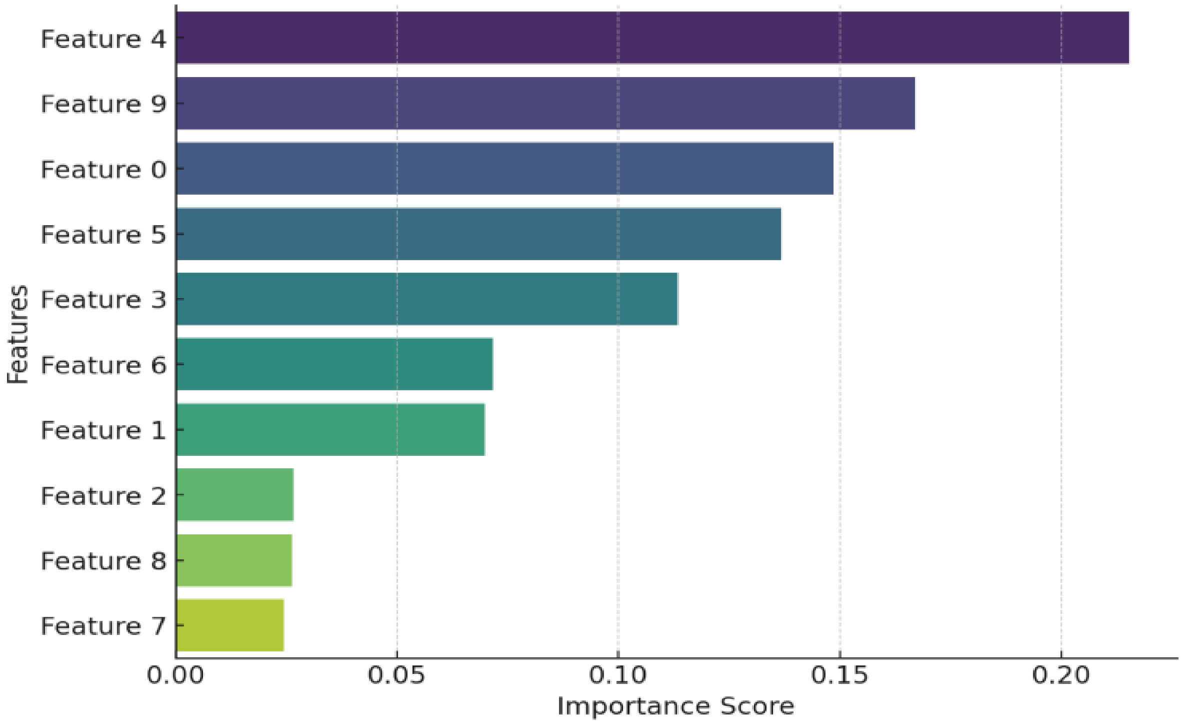

Fig. 3. Feature Importance Plot for Random Forest Model

Furthermore, the feature importance plot in Figure 3 demonstrates the relative contribution of each predictor variable in the Random Forest model. Environmental temperature, vaccination history, and flock age emerged as the most influential features, underscoring the importance of integrating both clinical and environmental data in predictive poultry health models.

Reference Tables 1 and 2: Among the tested algorithms, the Random Forest model demonstrated the highest predictive accuracy for fowl cholera infection status. It achieved an accuracy of 94.6%, with a precision of 93.8%, a recall of 95.2%, and an impressive ROC-AUC score of 0.97. Gradient Boosting performed slightly lower, attaining an accuracy of 93.2%, while Logistic Regression recorded the lowest accuracy at 89.5%. These results indicate that ensemble-based approaches, particularly Random Forest, were more effective in capturing the complex patterns in the dataset compared to linear models.

Analysis of feature importance from the Random Forest model revealed that environmental sanitation score was the most influential predictor of fowl cholera infection status. This was followed by vaccination history and the age of birds. Additional key predictors included mortality rate trends and respiratory symptoms, both of which contributed significantly to the classification performance. These findings highlight the role of farm hygiene practices, preventive vaccination programs, and observable clinical signs in predicting disease occurrence.

DISCUSSION

The findings underscore the potential of combining machine learning with veterinary epidemiology to overcome diagnostic limitations in fowl cholera management. The Random Forest model’s high performance confirms the robustness of ensemble learning in handling nonlinear relationships and mixed data types. Environmental sanitation and vaccination history emerged as leading predictors, consistent with global findings on biosecurity and preventive immunization [2], [1].The study’s alignment with precision agriculture principles indicates its potential for integration into smart farm monitoring systems, allowing near real-time alerts to farmers. However, limitations must be acknowledged. The dataset’s modest size (n = 500) and geographical restriction to Osun State reduce generalizability. The lack of an independent test set also limits external validation. Future expansions across diverse ecological and management contexts are necessary to ensure broader applicability.

Moreover, while three algorithms were tested, including XGBoost or deep learning architectures could enhance feature learning and model interpretability in future research. A clearer articulation of preprocessing and feature selection steps, as demonstrated in this revised methodology, contributes to methodological transparency and reproducibility.

CONCLUSION

This study demonstrates that advanced data mining techniques can significantly improve diagnostic accuracy for fowl cholera prediction. The Random Forest model, with its strong predictive metrics, proved most effective, emphasizing that farm sanitation, vaccination compliance, and bird age are crucial determinants of infection risk. By bridging veterinary science and data analytics, this research contributes a scalable, data-driven diagnostic framework for poultry disease surveillance one that could be embedded within digital platforms for early detection and response.

Future Work

Future research should integrate IoT-based real-time sensors for continuous monitoring of temperature, humidity, and air quality. These data streams can dynamically feed predictive models, allowing early warnings before disease outbreaks. The creation of mobile-based decision-support tools will empower farmers with actionable insights and automated alerts. Expanding datasets across regions will enhance generalization, accommodating differences in strain ecology and management systems.

Finally, the inclusion of deep learning for instance, image-based symptom detection using camera feeds offers a transformative opportunity to identify behavioral and visual indicators of fowl cholera early, strengthening predictive accuracy and timely intervention.

REFERENCES

1. 1 J. P. Christensen and M. Bisgaard, “Fowl cholera: Pathogenesis, diagnosis, and control,” Avian Pathology, vol. 51, no. 2, pp. 123–139, 2022. doi: 10.1080/03079457.2021.2001234.
2. 2 A. O. Ogunleye and A. T. Ajuwape, “Epidemiology and control of fowl cholera in Nigeria,” Nigerian Veterinary Journal, vol. 42, no. 3, pp. 145–154, 2021. doi: 10.4314/nvj.v42i3.6.
3. 3 L. Zhao, Y. Li, and Z. Wang, “Predictive modeling for poultry diseases using environmental and clinical data,” Veterinary Sciences, vol. 7, no. 4, p. 205, 2020. doi: 10.3390/vetsci7040205.
4. 4 S. Ahmed, M. Rahman, and A. Karim, “Application of machine learning in livestock disease diagnosis: A review,” Computers and Electronics in Agriculture, vol. 210, p. 107914, 2023. doi: 10.1016/j.compag.2023.107914.
5. 5 WOAH (World Organisation for Animal Health), Fowl cholera (Chapter 3.3.9), in WOAH Terrestrial Manual 2021, pp. 1–15.
6. 6 S. N. Shah, I. I. Ali, M. Yasir, M. Izhar, R. M. Tahir, A. Nadeem, and M. Zulfiqar, “The public health importance and management of infectious poultry diseases in smallholder systems: A review,” Veterinary Medicine International, vol. 2024, Art. no. 8838026, 2024. doi: 10.1155/2024/8838026.
7. 7 L. Ouyang, M. R. Campler, S. Wong, N. Xiao, and A. G. Arruda, “Exploring the impact of land cover on the occurrence of ornithobacteriosis and fowl cholera: A case-case study,” Animals, vol. 15, no. 3, p. 396, 2025. doi: 10.3390/ani15030396.
8. 8 F. G. Kebede, M. Birhanu, D. Boit, and A. Regassa, “Challenges to the poultry industry: Current perspectives and strategic approaches to enhance health and productivity,” Frontiers in Veterinary Science, vol. 7, p. 516, 2020. doi: 10.3389/fvets.2020.00516.
9. 9 Z. A. Tasew, D. B. Assefa, and M. S. Mulatu, “Fowl cholera in chickens: Trends in diagnosis and antimicrobial resistance in Gondar City, Ethiopia,” Veterinary Medicine International, vol. 2024, Art. no. 5592269, 2024. doi: 10.1155/2024/5592269.
10. 10 X. Zou, et al., “Epidemiological study of Newcastle disease in chicken farms in China (2019–2022) with a random-forest risk model,” Frontiers in Veterinary Science, vol. 11, Art. no. 1385353, 2024. doi: 10.3389/fvets.2024.1385353.
11. 11 A. Steiner, et al., “Machine learning frameworks to prioritize disease surveillance: Lessons from animal welfare inspections,” Computers in Biology and Medicine, vol. 183, p. 106013, 2024. doi: 10.1016/j.compbiomed.2024.106013.
12. 12 S. Mubarik, et al., “From data to diagnosis: Machine learning revolutionizes infectious disease detection and prediction,” Information, vol. 15, no. 11, p. 719, 2024. doi: 10.3390/info15110719.
13. 13 T. Zheng, et al., “Development and validation of machine-learning models for behavior detection in poultry,” Poultry Science, vol. 103, no. 10, Art. no. 103256, 2024. doi: 10.1016/j.psj.2024.103256.
14. 14 B. Martínez-López, J. M. Díaz-Cao, M. J. Clavijo, C. González-Crespo, and X. Liu, “Toward precision veterinary epidemiology: Applications, challenges, and opportunities of digitalization and the big-data revolution in livestock health,” Journal of the American Veterinary Medical Association, early access, 2025. doi: 10.2460/javma.25.01.0026.
15. 15 Z. Abidin, et al., “Exposure variables in veterinary epidemiology: Are they telling us what we think?” Frontiers in Veterinary Science, vol. 11, Art. no. 1442308, 2024. doi: 10.3389/fvets.2024.1442308.