A Generalizable AI-Driven Decision Support System for Infectious Disease Modeling

Marzieh Soltani[†,*], Rozita Dara[†], Shayan Sharif[‡]
[†] School of Computer Science, University of Guelph, Guelph, Ontario, Canada
[‡] Department of Pathobiology, University of Guelph, Guelph, Ontario, Canada

Abstract

The increasing complexity of infectious disease dynamics highlights the need for integrated and proactive surveillance systems. Traditional approaches remain largely reactive, relying on confirmed case reports and lagging indicators. This work presents a generalizable AI-driven Decision Support System (DSS) for spatiotemporal disease modeling, developed and validated using avian influenza surveillance in Canada. The proposed DSS consists of three core components: a digital surveillance module that leverages online activity for early warning signals; a spatiotemporal risk prediction module that models geographic disease risk using multi-source environmental and ecological data; and an expert system dashboard that integrates analytical outputs into an interactive, user-centered interface. The proposed DSS aims to equip policymakers and emergency responders with the tools needed to mitigate the impact of AIV outbreaks, through more informed, timely, and targeted interventions.

Keywords: Decision Support System, Disease Modeling, Avian Influenza, Spatiotemporal Surveillance, Social Media Analytics, Public Health AI

1. Introduction

Decision Support Systems (DSS) are interactive, data-driven tools that help decision-makers address complex and uncertain problems by integrating data management, analytical models, and domain knowledge to support public health surveillance, risk assessment, and intervention planning [1]. The growing threat of infectious diseases, particularly zoonotic spillovers, poses major challenges to public health systems, with Avian Influenza Virus (AIV) representing a highly dynamic threat due to rapid mutation and cross-species transmission [2, 3]. In recent years, highly pathogenic avian influenza outbreaks have caused significant economic losses in poultry production and raised serious public health concerns because of their zoonotic potential [4–6]. Human infections with subtypes such as H5N1 and H7N9 further highlight the ongoing pandemic risk [7–9], emphasizing the need for integrated predictive decision support tools that enable early detection and proactive disease management.

Traditional surveillance systems rely primarily on laboratory-confirmed cases, veterinary inspections, and formal reporting mechanisms. They depend on lagging indicators such as confirmed cases and laboratory reports, which often reflect outbreaks only after transmission has progressed. Epidemic intelligence approaches have incorporated informal sources and early event detection [10, 11]. Early Internet-based systems such as ProMED-mail and related platforms demonstrated the feasibility of mining online information for emerging threats [12, 13]. Despite these advances, many systems remain fragmented or lack structured integration with predictive analytics and operational workflows.

Recent advances in Artificial Intelligence (AI), natural language processing, and large-scale data analytics provide new opportunities to enhance infectious disease DSS design [14]. Modern deep learning approaches, particularly Large Language Models (LLMs) [15], enable the analysis of large-scale and heterogeneous datasets, including textual, epidemiological, and contextual data. In the context of avian influenza, prior research has identified environmental, ecological, and trade-related factors that influence virus spread [16, 17]. At the

[*] soltanik@uoguelph.ca

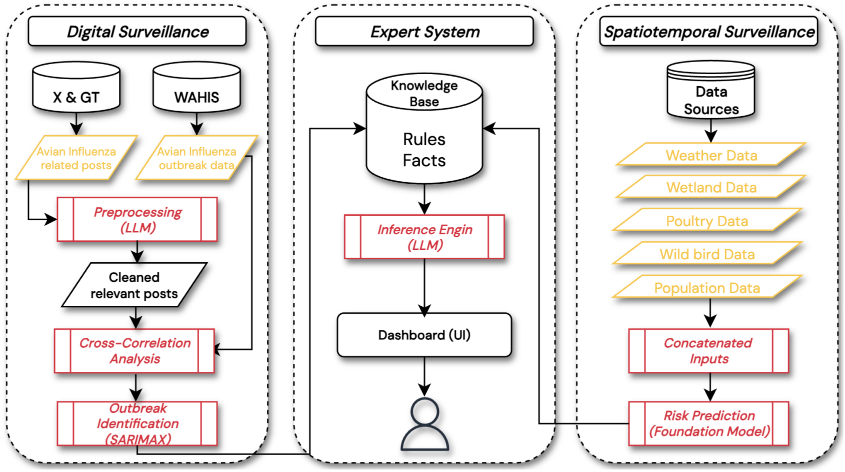

Figure 1. A high-level schematic representation of the architecture of the proposed DSS.

same time, digital platforms such as social media and search engines have shown potential for early outbreak detection across multiple diseases [18, 19]. Despite these advances, many systems analyze these data sources separately. Integrating digital surveillance, spatiotemporal risk monitoring, and knowledge-based decision logic within a unified DSS can provide a more comprehensive and timely approach to outbreak monitoring.

The primary objective of this study is to design and validate a generalizable AI-driven DSS framework for infectious disease modeling, using avian influenza surveillance in Canada as a case study. The framework includes three components: digital surveillance using social media and search trends, spatiotemporal surveillance that contextualizes outbreaks across space and time, and an expert system layer that converts analytical outputs into structured decision support. The main contribution of this work is a unified AI-driven DSS that integrates digital and geospatial data into a single operational framework, combining structured digital signal extraction, spatial representation learning using multi-channel tensors with pseudo-absence generation, and an interactive decision-support interface for transforming complex AI outputs into actionable insights for proactive disease mitigation by public health officials and epidemiologists.

2. AI-Driven DSS Framework

2.1. Overall Architecture

Traditional disease surveillance systems, particularly in avian influenza modeling, are limited by reliance on lagging indicators and restricted data sources [20, 21]. To address these gaps, we propose a unified DSS composed of three core components, as illustrated in Figure 1. The first is a digital surveillance module that leverages social media discourse and search engine trends to provide early warning signals that complements reactive approaches. The second component is spatiotemporal surveillance, which integrates epidemiological, environmental, and ecological information within a unified framework to support regional risk prediction and identify evolving spatial patterns. The third component is an expert system implemented as an interactive dashboard that integrates outputs from both modules into a single interface, connecting digital early warning signals with predictive spatial risk maps. The temporal forecasting and spatial risk models operate independently, with their outputs integrated into a knowledge base and used by an LLM-based module to support practical public health decision-making.

2.2. Methodological Foundations

Auto-regressive time series forecasting remains a cornerstone of outbreak prediction as it links historical case counts to future trends [22, 23]. However, these models are inherently reactive since they rely on lagged clinical data, which can delay public health responses. The proposed framework extends this approach by incorporating responsive digital indicators, such as search trends and social media activity, as exogenous inputs [24]. By capturing real-time shifts in public attention, these signals improve predictive performance and provide earlier indications for more proactive interventions. To maintain signal quality, LLMs are used to filter and identify relevant social media posts, enabling digital indicators to be integrated into forecasting pipelines in a structured and reliable way. For spatiotemporal monitoring, we used a foundation model, a transformer that integrates environmental, ecological, and epidemiological data linked to AIV [25–27] as multi-channel spatial inputs, allowing the model to capture both local and long-range spatial patterns. Foundation models pretrained on large datasets can then be adapted to outbreak modeling with limited additional training, allowing the system to generalize across regions where labeled outbreak data are limited.

2.3. Dashboard and End-User Engagement

The practical impact of a predictive AI framework in public health depends on how accessible it is to decision-makers. To translate complex model outputs into actionable insights, the DSS includes an expert system with an interactive dashboard that integrates results from the digital early warning and spatiotemporal modules in a clear interface. The system combines surveillance signals and modeling outputs into a unified knowledge base, where LLM-driven reasoning helps interpret and synthesize rules while providing transparent explanations of alerts so users can understand the evidence behind them. The dashboard supports both descriptive and predictive analytics through time-series views of digital signals and layered spatial risk summaries, and its design is guided by real end-user workflows with iterative feedback from epidemiologists and public health policymakers to ensure the system improves situational awareness and supports timely, evidence-based decisions.

3. Case Study: Avian Influenza Surveillance in Canada

3.1. Digital Early Warning System

Our analysis of digital indicators showed that internet search queries and social media activity are significantly correlated with the temporal progression of AIV outbreaks [24]. Cross-correlation analysis identified statistically significant associations across geographic scales, with digital signals preceding officially reported case counts by one to two weeks. Both Google Trends and X activity showed positive correlations at negative lags, indicating increased online engagement before formal outbreak confirmation. These results suggest that public attention and online information-seeking behavior respond quickly to emerging outbreaks and can serve as early indicators of disease spread.

To quantify these temporal associations, we implemented a relevance classification pipeline to filter outbreak-related posts from large volumes of social media data. Among several transformer-based models, DistilBERT [28, 29], achieved the best performance with mean precision, recall, and F1-score of 84.65%, 87.25%, and 85.91% using 5-fold cross-validation. The refined digital signals were then analyzed using cross-correlation against confirmed outbreak counts, revealing consistent lead times of one to two weeks across geographic scales, suggesting that digital signals can act as early indicators of outbreak dynamics. The filtered indicators were incorporated as exogenous variables into Seasonal AutoRegressive Integrated

Table 1. SARIMAX Performance on Global Data with and without Lagged Exogenous Variables. The *Best Orders* column represents the selected **(p, d, q)** order and seasonal **(P, D, Q, s)** order for each model configuration.

Scenario	**X**	**GT**	R^2 **Score**	**Best Orders**	**Scenario**	**X**	**GT**	R^2 **Score**	**Best Orders**
Exog. without lag (Train: 52w, Eval: 13w)					*Exog. -3 weeks lag (Train: 49w, Eval: 13w)*				
(i)			0.2032 ± 0.0031	$(1,1,1)(0,0,2,52)$	(i)			$0.1262 \pm 2.78\text{e}{-17}$	$(1,1,1)(2,0,0,52)$
(ii)	✓		0.2219 ± 0.0035	$(2,1,0)(1,0,0,52)$	(ii)	✓		$0.4024 \pm 5.55\text{e}{-17}$	$(2,1,2)(0,0,0,52)$
(iii)		✓	0.2270 ± 0.0038	$(1,1,1)(0,0,1,52)$	(iii)		✓	$0.1953 \pm 5.55\text{e}{-17}$	$(0,0,0)(2,0,2,52)$
(iv)	✓	✓	0.2491 ± 0.0027	$(1,1,1)(0,0,1,52)$	(iv)	✓	✓	$\mathbf{0.4176 \pm 5.55e{-17}}$	$(1,1,2)(0,0,0,52)$

Moving Average with eXogenous variables (SARIMAX) models. An ablation study compared endogenous-only configurations with models including social media activity, search trends, and lagged variants. As shown in Table 1, models without lagged exogenous variables provided moderate improvements over baseline autoregressive models, but introducing a three-week lag informed by cross-correlation notably improved performance. Each experiment was run 30 times and evaluated using the average R^2 score, a scale-independent regression metric ranging between $(-\infty, 1]$. The combined model using lagged X activity and Google Trends achieved an R^2 of 0.4176 compared to 0.1262 for the endogenous-only model, demonstrating that properly aligned digital signals improve forecasting accuracy and provide earlier warning beyond traditional case-based models.

To better understand public response during outbreak periods, we analyzed social media discourse using LLM-based approaches, including DistilBERT [28], Mixtral-8x7B [30], BERTopic [31], and RoBERTa [32], for filtering, topic modeling, sentiment, and emotion analysis. The results showed dominant negative sentiments such as fear, anger, and sadness, along with increased risk-focused and policy-related discussions during major outbreak waves [33]. These findings indicate that digital platforms provide both early warning signals and insight into public perception, supporting their role within the DSS framework.

3.2. **Geospatial Risk Prediction**

For the spatiotemporal component, we focused on integrating multiple data sources that have been identified in the avian influenza literature as relevant to virus spread and persistence. These include bioclimatic and weather-related variables, wetland distribution, poultry density, wild bird observations, and human population density [26, 34–38]. These factors have been consistently associated with AIV transmission in epidemiological and ecological studies. The objective was to combine these diverse spatial data sources within a unified modeling framework that learns joint patterns across them. Because confirmed outbreak records are usually available only for positive cases, supervised learning is supported through background sampling (pseudo-absence generation), enabling the model to distinguish high-risk regions from non-outbreak areas. This component therefore provides spatial risk indicators that complement digital surveillance and support comprehensive disease monitoring.

4. **Conclusion**

The primary contribution of this research is a unified AI-driven DSS that integrates digital epidemiology and geospatial risk monitoring to support proactive disease surveillance. This approach addresses the limitations of reactive systems by enabling public health officials to track temporal trends through internet-based early warning signals while assessing geographic risk patterns in parallel. The modular design allows the framework to adapt across diseases and regions by incorporating new digital indicators and environmental data without major restructuring. Ongoing work focuses on validating and refining the system through user studies, and on exploring uncertainty estimation in the spatial component to better communicate confidence levels in public health decisions.

References

[1] B. Almadani, H. Kaisar, I. R. Thoker, and F. Aliyu. "A systematic survey of distributed decision support systems in healthcare". In: *Systems* 13.3 (2025), p. 157.

[2] R. N. Shanta, M. Akther, M. A. Prodhan, S. H. Akter, H. Annandale, S. Sarker, S. Abraham, and J. M. Uddin. "Adaptation and Outbreak of Highly Pathogenic Avian Influenza in Dairy Cattle: An Emerging Threat to Humans, Pets, and Peridomestic Animals". In: *Pathogens* 14.9 (2025), p. 846.

[3] J. Charostad, M. R. Z. Rukerd, S. Mahmoudvand, D. Bashash, S. M. A. Hashemi, M. Nakhaie, and K. Zandi. "A comprehensive review of highly pathogenic avian influenza (HPAI) H5N1: An imminent threat at doorstep". In: *Travel medicine and infectious disease* 55 (2023), p. 102638.

[4] R. M. Seeger, A. D. Hagerman, K. K. Johnson, D. L. Pendell, and T. L. Marsh. "When poultry take a sick leave: Response costs for the 2014–2015 highly pathogenic avian influenza epidemic in the USA". In: *Food Policy* 102 (2021), p. 102068.

[5] K. K. Johnson, R. M. Seeger, and T. L. Marsh. "Local economies and highly pathogenic avian influenza". In: *Choices* 31.2 (2016), pp. 1–9.

[6] R. Xie, K. M. Edwards, M. Wille, X. Wei, S.-S. Wong, M. Zanin, R. El-Shesheny, M. Ducatez, L. L. Poon, G. Kayali, et al. "The episodic resurgence of highly pathogenic avian influenza H5 virus". In: *Nature* (2023), pp. 1–8.

[7] R. Yang, H. Sun, F. Gao, K. Luo, Z. Huang, Q. Tong, H. Song, Q. Han, J. Liu, Y. Lan, et al. "Human infection of avian influenza A H3N8 virus and the viral origins: A descriptive study". In: *The Lancet Microbe* 3.11 (2022), e824–e834.

[8] U. AbuBakar, L. Amrani, F. A. Kamarulzaman, S. A. Karsani, P. Hassandarvish, and J. E. Khairat. "Avian influenza virus tropism in humans". In: *Viruses* 15.4 (2023), p. 833.

[9] D. A. Philippon, P. Wu, B. J. Cowling, and E. H. Lau. "Avian influenza human infections at the human-animal interface". In: *The Journal of Infectious Diseases* 222.4 (2020), pp. 528–537.

[10] C. Paquet, D. Coulombier, R. Kaiser, and M. Ciotti. "Epidemic intelligence: a new framework for strengthening disease surveillance in Europe". In: *Eurosurveillance* 11.12 (2006), pp. 5–6.

[11] E. Christaki. "New technologies in predicting, preventing and controlling emerging infectious diseases". In: *Virulence* 6.6 (2015), pp. 558–565.

[12] V. L. Yu and L. C. Madoff. "ProMED-mail: an early warning system for emerging diseases". In: *Clinical infectious diseases* 39.2 (2004), pp. 227–232.

[13] J. S. Brownstein and C. Freifeld. "HealthMap: the development of automated real-time internet surveillance for epidemic intelligence". In: *Weekly releases (1997–2007)* 12.48 (2007), p. 3322.

[14] S. A. Alowais, S. S. Alghamdi, N. Alsuhebany, T. Alqahtani, A. I. Alshaya, S. N. Almohareb, A. Aldairem, M. Alrashed, K. Bin Saleh, H. A. Badreldin, et al. "Revolutionizing healthcare: the role of artificial intelligence in clinical practice". In: *BMC medical education* 23.1 (2023), p. 689.

[15] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, et al. "A survey of large language models". In: *arXiv preprint arXiv:2303.18223* 1.2 (2023), pp. 1–124.

[16] K. A. Herrick, F. Huettmann, and M. A. Lindgren. "A global model of avian influenza prediction in wild birds: the importance of northern regions". In: *Veterinary research* 44 (2013), pp. 1–9.

[17] G. Fournié, F. Guitian, P Mangtani, and A. Ghani. "Impact of the implementation of rest days in live bird markets on the dynamics of H5N1 highly pathogenic avian influenza". In: *Journal of the Royal Society Interface* 8.61 (2011), pp. 1079–1089.

[18] M. J. Paul, M. Dredze, and D. Broniatowski. "Twitter improves influenza forecasting". In: *PLoS currents* 6 (2014).

[19] S. Yousefinaghani, R. Dara, S. Mubareka, and S. Sharif. "Prediction of COVID-19 waves using social media and Google search: a case study of the US and Canada". In: *Frontiers in public health* 9 (2021), p. 656635.

[20] C. Azat, M. Alvarado-Rybak, J. F. Aguilera, and J. A. Benavides. "Spatio-temporal dynamics and drivers of highly pathogenic avian influenza H5N1 in Chile". In: *Frontiers in Veterinary Science* 11 (2024), p. 1387040.

[21] C. Duan, C. Li, R. Ren, W. Bai, and L. Zhou. "An overview of avian influenza surveillance strategies and modes". In: *Science in One Health* (2023), p. 100043.

[22] M. J. Kane, N. Price, M. Scotch, and P. Rabinowitz. "Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks". In: *BMC bioinformatics* 15 (2014), pp. 1–9.

[23] L. J. Kjær, M. P. Ward, A. E. Boklund, L. E. Larsen, C. K. Hjulsager, and C. T. Kirkeby. "Using surveillance data for early warning modelling of highly pathogenic avian influenza in Europe reveals a seasonal shift in transmission, 2016–2022". In: *Scientific reports* 13 (2023).

[24] Authors. "Leveraging social media and google trends to identify waves of avian influenza outbreaks in USA and Canada". In: *Expert Systems with Applications* 291 (2025), p. 128482.

[25] S. Hayes, J. Hilton, J. Mould-Quevedo, C. A. Donnelly, M. Baylis, and L. Brierley. "Ecology and environment predict spatially stratified risk of H5 highly pathogenic avian influenza clade 2.3. 4.4 b in wild birds across Europe". In: *Scientific Reports* (2025).

[26] M.-C. Dupas, M. F. Vincenti-Gonzalez, M. Dhingra, C. Guinat, T. Vergne, W. Wint, G. Hendrickx, C. Marsboom, M. Gilbert, and S. Dellicour. "Global risk mapping of highly pathogenic avian influenza H5N1 and H5Nx in the light of epidemic episodes occurring from 2020 onward". In: (2025). DOI: 10.7554/elife.104748.3. URL: http://dx.doi.org/10.7554/eLife.104748.3.

[27] S. Yousefinaghani, R. A. Dara, Z. Poljak, and S. Sharif. "A decision support framework for prediction of avian influenza". In: *Scientific Reports* 10.1 (2020), p. 19011.

[28] V Sanh. "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter". In: *arXiv preprint arXiv:1910.01108* (2019).

[29] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805* (2018).

[30] A. Q. Jiang, A. Sablayrolles, A. Roux, A. Mensch, B. Savary, C. Bamford, D. S. Chaplot, D. d. l. Casas, E. B. Hanna, F. Bressand, et al. "Mixtral of experts". In: *arXiv preprint arXiv:2401.04088* (2024).

[31] M. Grootendorst. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure". In: *arXiv preprint arXiv:2203.05794* (2022).

[32] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. "Roberta: A robustly optimized bert pretraining approach". In: *arXiv preprint arXiv:1907.11692* (2019).

[33] Authors. "Analyzing social media discourse of avian influenza outbreaks". In: *Natural Language Processing Journal* (2025), p. 100176.

[34] S. K. Puttahonnappa, J. Anandakumar, N. N. Barman, R. Rajkumar, K. Paramanandham, S. S. Patil, S. Lamba, A. V. Patil, and B. R. Gulati. "Investigating environmental determinants and spatiotemporal dynamics of highly pathogenic avian influenza H5N1 outbreaks in India through machine learning". In: *Scientific Reports* 15.1 (2025), p. 36132.

[35] J. Schreuder, H. J. de Knegt, F. C. Velkers, A. R. Elbers, J. Stahl, R. Slaterus, J. A. Stegeman, and W. F. de Boer. "Wild bird densities and landscape variables predict spatial patterns in HPAI outbreak risk across The Netherlands". In: *Pathogens* 11.5 (2022), p. 549.

[36] M. R. Opata, A. Lavarello-Schettini, J. C. Semenza, and J. Rocklöv. "Predictiveness and drivers of highly pathogenic avian influenza outbreaks in Europe". In: *Scientific Reports* 15.1 (2025), p. 20286.

[37] M. S. Dhingra, J. Artois, T. P. Robinson, C. Linard, C. Chaiban, I. Xenarios, R. Engler, R. Liechti, D. Kuznetsov, X. Xiao, et al. "Global mapping of highly pathogenic avian influenza H5N1 and H5Nx clade 2.3. 4.4 viruses with spatial cross-validation". In: *elife* 5 (2016), e19571.

[38] J. Belkhiria, R. J. Hijmans, W. Boyce, B. M. Crossley, and B. Martínez-López. "Identification of high risk areas for avian influenza outbreaks in California using disease distribution models". In: *PLoS One* 13.1 (2018), e0190824.