Journal of Poultry Sciences and Avian Diseases

Journal homepage: www.jpsad.com



Machine learning-driven identification of immune signatures from RNA-Seq data in H5N1-infected chickens: a computational Immunology approach

Fatemeh Keivan[1], Gholamreza Nikbakht Brujeni[1*]

1. Department of Microbiology and Immunology, Faculty of Veterinary Medicine, University of Tehran, Tehran, Iran

*** Corresponding author email address:** nikbakht@ut.ac.ir

A r t i c l e I n f o

Article type:
Original Research

How to cite this article:
Keivan, F. & Nikbakht Brujeni, G. (2026). Machine learning-driven identification of immune signatures from RNA-Seq data in H5N1-infected chickens: a computational Immunology approach. *Journal of Poultry Sciences and Avian Diseases, 4*(4), 1-17.
http://dx.doi.org/10.61838/kman.jpsad.217

© 2026 the authors. Published by SANA Institute for Avian Health and Diseases Research, Tehran, Iran. This is an open access article under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License.

A B S T R A C T

Highly Pathogenic Avian Influenza (H5N1) continues to pose a serious risk to public health and poultry. In order to differentiate H5N1-infected from healthy chicken samples and to guide the development of diagnostics and vaccines, we postulated that Machine Learning (ML) applied to RNA sequencing (RNA-Seq) data could biologically detect meaningful immune gene signatures. While previous transcriptomic studies have characterized host responses to H5N1 infection in chickens, the application of interpretable machine-learning approaches to prioritize immune-associated transcriptional signatures from chicken RNA-Seq datasets remains relatively unexplored. RNA-Seq data from H5N1-infected chicken lung and ileum tissues (ArrayExpress E-MTAB-2908) that were publicly available, were chosen. Fivefold stratified cross-validation (~80/20 train/test per fold) was used to train two supervised ML models, such as Random Forest (RF) and linear-kernel Support Vector Machine (SVM). Performance was evaluated using Area Under Curve (AUC) and Receiver Operating Characteristic **(**ROC) curves. RF reached a mean AUC=0.85, while SVM-Linear achieved AUC=0.75. Top-ranking interferon-stimulated genes (ISGs), including IFIT5, MX1, and OASL, were consistently upregulated in infected samples, indicating activation of type I interferon pathways. Concordant findings across models support the stability and biological relevance of the identified signatures despite the modest sample size. These findings demonstrate that interpretable ML approaches can successfully prioritize biologically relevant antiviral signatures from chicken RNA-Seq datasets. However, the identified signatures should be considered candidate computational immune signatures that require independent experimental validation before potential application in biomarker development, disease surveillance, or vaccine-related research.

Keywords: *H5N1, RNA-Seq, Machine Learning (ML), Chicken Immune Response, Immunoinformatics.*

Article history:
Received 4 June 2025
Revised 13 July 2025
Accepted 29 July 2025
Published online 01 October 2026

1 Introduction

One of the most serious viral threats to poultry is still Highly Pathogenic Avian Influenza (HPAI) A virus pathotype H5N1, which can lead to significant financial losses and human zoonotic risks (Rehman et al., 2022). The causative agent is an enveloped, negative-sense single-stranded RNA virus belonging to the Orthomyxoviridae family, with a segmented genome that facilitates genetic reassortment and evolution (Ding et al., 2025). The molecular mechanisms behind the chicken's immune response to H5N1 infection remain unclear despite a great deal of research (Morris et al., 2023). Chickens are highly susceptible to H5N1, whereas other avian species, such as ducks, are relatively resistant, suggesting species-specific differences in innate immune signaling that remain poorly understood (Evseev & Magor, 2019).

Macrophages and other innate immune cells, as first-line defenders against viral invasion, play a critical role in sensing viral RNA through pattern recognition receptors such as MDA5 and related cytosolic viral RNA sensors, triggering downstream type I interferon responses (Liniger et al., 2012). Beyond viral sensing, these cells also contribute to innate immunity through phagocytosis of viral particles, secretion of pro-inflammatory cytokines (e.g., IL-1β, IL-6, TNF-α), and antigen presentation to adaptive immune cells (Hassan & Sharif, 2025). In the context of H5N1 infection, innate immune cells, including macrophages, play important roles during H5N1 infection, particularly within respiratory tissues (Perrone et al., 2008).

The antiviral and inflammatory responses initiated by immune cells during H5N1 infection Lead to widespread alterations in host gene expression, generating what are commonly referred to as immune signatures (Ahmed et al., 2016). During H5N1 infection, host tissues undergo extensive transcriptional reprogramming in response to viral recognition and interferon signaling (Golpasand et al., 2025). These transcriptional changes generate complex immune signatures composed of coordinated expression patterns across numerous antiviral and inflammatory genes (Kuchipudi et al., 2015).

Traditional approaches to studying host responses to H5N1 infection have relied on targeted methods such as quantitative PCR (qPCR), ELISA, and immunohistochemistry, which measure a limited set of pre-selected genes or proteins (Lowe et al., 2017). While these methods have provided valuable insights, they are inherently biased and cannot capture the full complexity of the host transcriptome (Stark et al., 2019). Unlike microarray-based approaches, RNA sequencing (RNA-Seq) enables unbiased transcriptome-wide profiling, allowing detection of both highly and lowly expressed genes, novel transcripts, and subtle immune-related expression changes during infection (Wang et al., 2009). This comprehensive coverage makes RNA-Seq particularly suitable for investigating complex host–pathogen interactions such as H5N1 infection (Westermann et al., 2017). Because such signatures arise from simultaneous interactions among many genes rather than isolated biomarkers, transcriptome-wide approaches such as RNA-Seq are required to capture the complexity of host immune responses accurately (Wang et al., 2009).

Traditional differential expression analysis (e.g., edgeR, DESeq2) focuses on individual genes in isolation and may overlook subtle, synergistic interactions among genes that collectively define complex immune phenotypes (Huang et al., 2017; Love et al., 2014). More robust and biologically interpretable analytical approaches that can identify important immune-related genes involved in host defense mechanisms are now required as a result of this limitation (Libbrecht & Noble, 2015). However, the enormous dimensionality of RNA-Seq datasets makes identification of biologically meaningful immune signatures extremely challenging using conventional statistical approaches alone (Conesa et al., 2016).

Artificial Intelligence (AI) and Machine Learning (ML) have increasingly become potent tools for biological data analysis in recent years, making it possible to find subtle multidimensional patterns in intricate transcriptomic datasets (Tarca, 2007). In Immunology and virology, ML models have been used to predict host-pathogen interactions, classify viral sequences, and identify transcriptional signatures associated with disease severity in human influenza and COVID-19 patients (Overmyer et al., 2021; Randhawa et al., 2020). However, there is still much to learn about their use in avian immunology, specifically in detecting immune signatures linked to H5N1 infection (Ahmed et al., 2016). Despite recent advances in transcriptomics and AI, the integration of ML with avian RNA-Seq datasets for immune signature detection in H5N1 infection remains largely unexplored (Ahmed et al., 2016; Libbrecht & Noble, 2015).

Previous avian ML studies have primarily focused on classifying viral subtypes from sequence data or predicting disease outcomes from clinical signs, with limited application to high-dimensional transcriptomic datasets for host gene signature discovery (Bzhalava et al., 2018; Shu et

al., 2016). Thus, this study hypothesizes that ML algorithms, namely Random Forest (RF) and Support Vector Machine (SVM), can efficiently detect reliable immune gene signatures that differentiate H5N1-infected from healthy chicken tissue samples.

AI-driven approaches are therefore uniquely positioned to uncover hidden multidimensional patterns within transcriptomic datasets (Libbrecht & Noble, 2015). Previous transcriptomic studies of H5N1 infection in chickens have primarily relied on differential gene expression and pathway-enrichment analyses to characterize host antiviral responses (Ahmed et al., 2016; Golpasand et al., 2025; Morris et al., 2023). In contrast, the application of interpretable ML approaches to prioritize immune-associated transcriptional signatures from chicken H5N1 RNA-Seq datasets has received limited attention. Therefore, the present study aimed to evaluate whether RF and SVM models could reliably identify biologically meaningful immune signatures associated with H5N1 infection using publicly available transcriptomic data.

This study represents a computational reanalysis of publicly available transcriptomic data and explores the utility of interpretable ML approaches for prioritizing biologically relevant immune signatures associated with H5N1 infection.

2 Material and Methods

The public repository ArrayExpress provided transcriptomic data from chicken lung and ileum tissues, including both infected and uninfected (control) groups, with accession number E-MTAB-2908 (https://www.omicsdi.org/dataset/atlas-experiments/E-MTAB-2908) (Smith et al., 2015). Transcriptomic gene-expression data were obtained from the E-MTAB-2908 dataset through the ArrayExpress repository. Raw gene-count matrices were used for downstream analyses, whereas raw FASTQ files were not reprocessed in the present study. Samples were subsequently labeled according to infection status (H5N1-infected or control) and imported into the machine-learning pipeline for preprocessing, normalization, and feature-selection procedures. A high-dimensional data structure appropriate for computational analysis was formed by the dataset, which comprised 24 samples (12 H5N1-infected and 12 control samples) with expression profiles of thousands of genes each. The dataset included lung and ileum tissues collected at 1and 3 days post-infection, with three biological replicates for each tissue and time-point combination. Each sample was labeled according to infection status (1=H5N1-infected, 0=control), and the expression values of selected genes served as input features for the models. Raw gene counts were log-transformed to stabilize variance and reduce the impact of extreme values in order to get the data ready for modeling. A pseudocount of 1 was added prior to log transformation, and genes with zero variance across samples were removed before feature selection. To reduce noise and improve the signal-to-noise ratio, genes exhibiting little variation in expression across samples were filtered out (Conesa et al., 2016). Details of all samples included in the analysis are provided in Supplementary Table S1. Only H5N1-infected and corresponding control samples from the E-MTAB-2908 dataset were included, while H5N2 samples were excluded from the analysis to avoid introducing subtype-related transcriptional variation that could confound the machine-learning classification task.

2.1 Feature selection and normalization

Raw gene-expression values were log2 transformed prior to downstream analysis to reduce data skewness and stabilize variance across genes. In addition, genes exhibiting very low variance across samples were removed because they were unlikely to contribute meaningful information to classification performance. An ANOVA F-test, which ranked genes based on their ability to discriminate between infected and control groups, was then used to select features (Pedregosa, 2011). Approximately 500 top-ranked genes were kept for further examination. The number of selected genes (n=500) was empirically chosen before model evaluation based on methodological considerations related to dimensionality reduction, model interpretability, and retention of potentially informative transcriptomic signals. Feature selection (ANOVA F-test) was performed within each training fold to prevent information leakage between training and validation data. Z-score normalization was used to standardize all variables in order to guarantee uniform feature scaling (Pedregosa, 2011).

2.2 Machine learning models

Two supervised ML algorithms were employed to classify samples and extract biologically meaningful patterns. The first model, RF, is a nonlinear ensemble technique that constructs multiple decision trees and combines their predictions to enhance accuracy and robustness (Breiman, 2001).

The RF classifier was implemented using 700 decision trees (n_estimators=700), with max_features set to sqrt, class_weight set to balanced, and random_state=42. These parameters were selected to provide robust performance while reducing the risk of overfitting in the high-dimensional transcriptomic dataset. The remaining parameters were retained at their default scikit-learn settings, including max_depth=None, min_samples_split=2, and min_samples_leaf=1.

The second model, SVM with a linear kernel (SVM-Linear), was chosen for its interpretability and ability to provide coefficient-based insights into gene relevance (Cortes & Vapnik, 1995).

The SVM model was implemented using a linear kernel with probability estimation enabled (probability=True), class_weight=balanced, and random_state=42. The linear kernel was selected for its interpretability and ability to provide coefficient-based estimates of feature importance. During model development, alternative regularization parameters (C=0.1, 1, 10, and 100, respectively) were evaluated using cross-validation, and the final SVM-Linear configuration was selected based on classification performance. Feature coefficients were derived from the final fitted SVM-Linear model after feature selection and model training rather than being averaged across individual cross-validation folds. The standard SVM hinge-loss optimization framework was used for classification.

RF was selected because of its robustness to noisy high-dimensional biological data and its ability to estimate feature importance (Breiman, 2001), whereas SVM-Linear was chosen due to its strong performance in small-sample transcriptomic datasets and its interpretability through coefficient-based feature weighting (Statnikov et al., 2005).

2.3 *Model training and cross-validation*

Both models were trained and evaluated using five-fold stratified cross-validation to ensure balanced representation of the infected and control conditions (Pedregosa, 2011). Each sample was precisely used once for validation by dividing the dataset into training and testing subsets within each of the five cross-validation folds (Pedregosa, 2011). By minimizing overfitting, this method made it possible to estimate the model's generalization performance on unseen data with confidence (Pedregosa, 2011). In order to guarantee a consistent evaluation throughout all iterations, roughly 80% of the samples in each fold were used for training and 20% for testing.

Stratified K Fold was implemented with shuffle True and random_state=42 to ensure reproducible partitioning of samples across the cross-validation folds. To assess feature-selection stability, the frequency with which key machine-learning-prioritized genes were selected across the five cross-validation folds was additionally evaluated. Selection consistency was summarized for the highest-ranked infection-associated genes identified by the RF and SVM-Linear models.

2.4 *Performance evaluation*

The model's ability to discriminate was evaluated using the AUC metric and Receiver Operating Characteristic (ROC) curves (Fawcett, 2006). The AUC metric was selected as it provides a threshold-independent measure of classification performance and is more reliable than accuracy for small, slightly imbalanced biological datasets (Fawcett, 2006).

2.5 *Interpretation and visualization*

The RF model's feature importance values were analyzed to determine the main immune-related genes that contribute to classification in order to interpret the biological significance of the predictions (Breiman, 2001). To illustrate the different expression patterns of these top-ranked genes between infected and control samples, a heatmap was created based on the genes identified by the RF model. In order to verify the consistency and reproducibility of the immune signatures found by the RF model, the SVM-Linear model was employed as a supplementary method (Breiman, 2001; Cortes & Vapnik, 1995). Principal Component Analysis (PCA) was additionally performed to visualize overall transcriptomic variation and sample distribution between infected and control groups based on normalized gene expression profiles (Ringnér, 2008). To further investigate potential sources of transcriptomic variation, additional PCA visualizations were generated using the normalized gene-expression matrix. Samples were colored according to infection status, tissue type, and time point to evaluate the relative contribution of these variables to global transcriptomic variation.

2.6 *Software and reproducibility*

Standard libraries like *scikit-learn* (version 1.5.1) for ML, *pandas* (version 2.2.2) and *NumPy* (version 1.26.4) for data manipulation, and *matplotlib* (version 3.9.0) and *seaborn*

(version 0.13.2) for visualization were used for all analyses, which were carried out in *Python* (version 3.10.5) (Mckinney, 2010; Pedregosa, 2011). No additional wet-lab experiments were performed, as the study was based entirely on publicly available RNA-Seq data. All scripts and analysis workflows were implemented in Python and are available upon request for reproducibility.

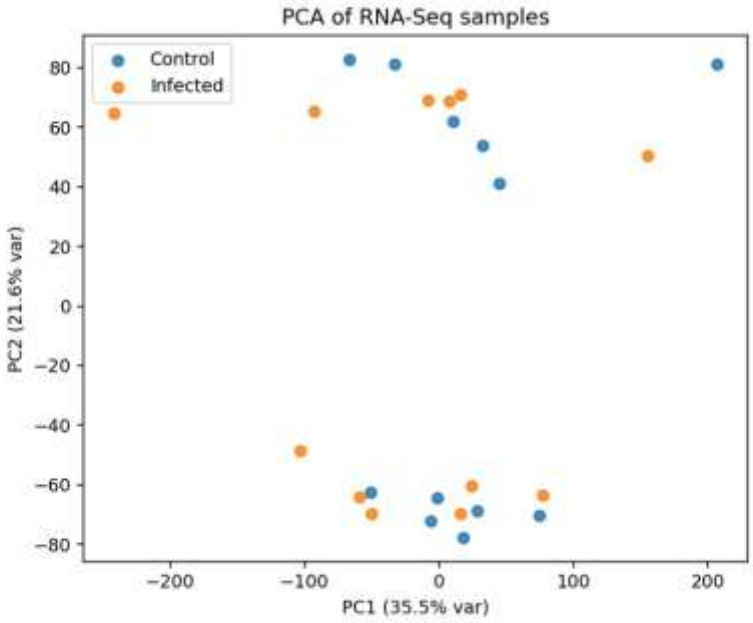
2.7 Differential-expression and Functional enrichment analysis

To further support the biological interpretation of the machine-learning findings, conventional differential-expression analysis was performed using DESeq2. Log2 fold-change values and adjusted *p*-values were calculated for ML-prioritized genes. In addition, functional annotation and pathway classification of machine-learning-prioritized genes were performed using Gene Ontology (GO) and Reactome pathway information available for the E-MTAB-2908 dataset. The resulting differential-expression profiles and functional categories were compared with genes prioritized by the machine-learning models and with known interferon-stimulated antiviral pathways to further assess their biological relevance.

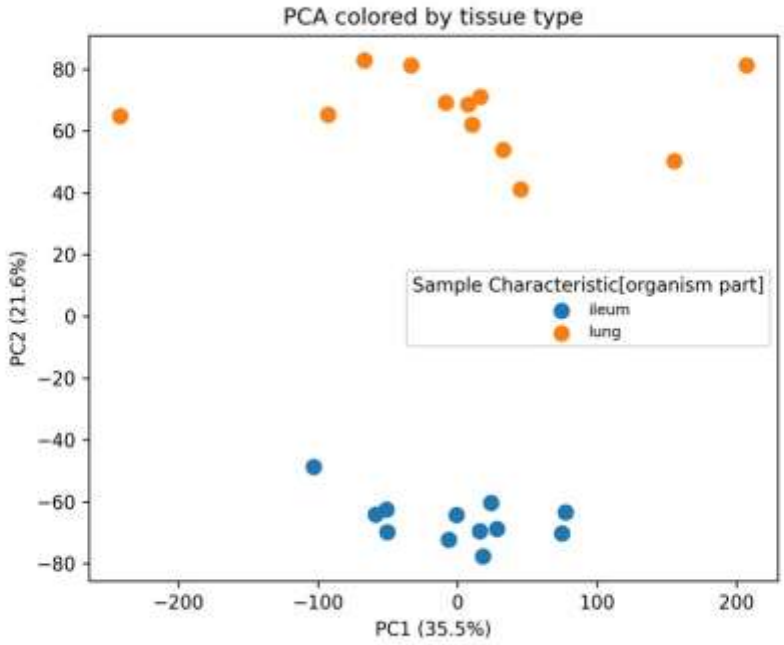
3 Results

3.1 PCA visualization of transcriptomic samples

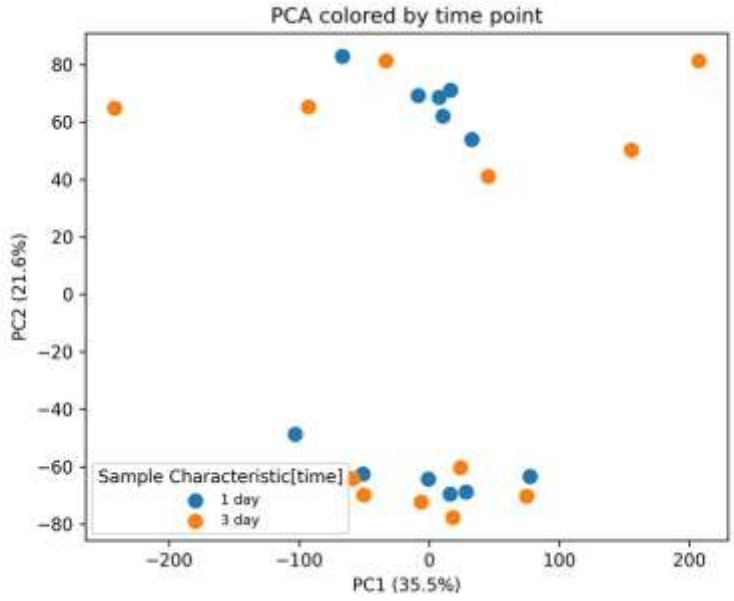
PCA was performed to visualize global transcriptomic variation between H5N1-infected and control chicken samples. Partial separation between infected and control groups was observed along the principal component axes, indicating infection-associated differences in gene expression profiles (Figure 1a). PC1 accounted for 35.5% of the total variance, whereas PC2 explained 21.6% of the variance. Although some overlap between groups remained, the observed clustering patterns suggested the presence of biologically relevant transcriptomic alterations associated with H5N1 infection. Additional PCA visualizations were generated to further examine transcriptomic variation according to tissue type and time point (Figure 1b and 1c). PCA plots colored by tissue type and time point are presented in Figure 1b and 1c, respectively. These visualizations provide complementary information regarding sample distribution across the first two principal components and potential sources of transcriptomic variation within the dataset.



a



b



c

Figure 1. PCA of chicken RNA-Seq transcriptomic samples. (a) PCA plot colored according to infection status (H5N1-infected versus control), showing partial separation between infected and control samples. (b) PCA plot colored according to tissue type (lung and ileum). (c) PCA plot colored according to time point (1 and3 days post-infection). PCA was performed using normalized gene-expression profiles to visualize major sources of transcriptomic variation within the dataset.

3.2 Classification performance

Both ML models demonstrated reliable classification performance in distinguishing H5N1-infected chicken samples from uninfected control samples. The RF model achieved the highest predictive performance, with a mean AUC value of 0.85 (Figure 2). The SVM-Linear model also demonstrated robust discrimination capability, achieving an AUC value of 0.75 (Figure 3). The classification performance of the RF and SVM-Linear models is summarized in Table 1. Besides ROC-AUC, additional evaluation metrics including accuracy, sensitivity, specificity, precision, and F1-score were calculated to provide a more comprehensive assessment of model performance. Detailed confusion matrices and fold-wise classification performance of RF and SVM-Linear models are presented in Table 2.

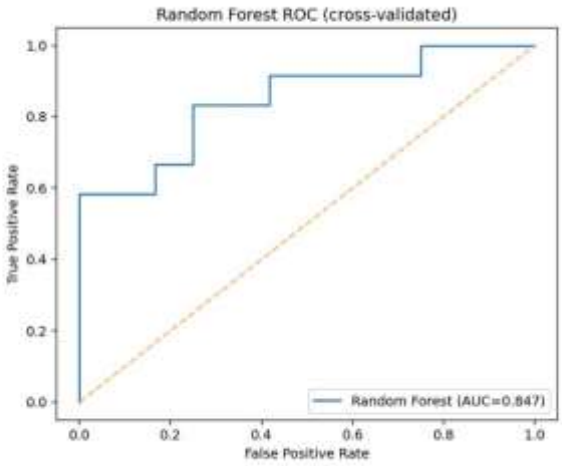


Figure 2. ROC curve of the RF model showing classification performance between infected and control samples (AUC=0.85).

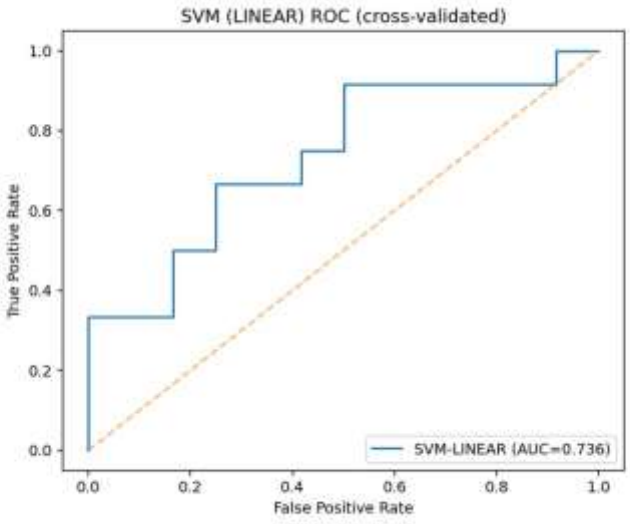


Figure 3. ROC curve of the SVM-Linear model showing classification performance between infected and control samples (AUC=0.75).

Across five-fold stratified cross-validation, both models consistently separated infected from control samples. Both RF and SVM-Linear models achieved reproducible classification performance despite the relatively limited sample size.

Table 1. Classification performance of RF and SVM-Linear models.

Model	AUC	Accuracy	Sensitivity (Recall)	Specificity	Precision	F1-score	Balanced Accuracy
RF	0.85	0.75	0.75	0.75	0.75	0.75	0.75
SVM-Linear	0.75	0.71	0.67	0.75	0.73	0.70	0.71

Table 2. Confusion matrices and fold-wise classification performance of RF and SVM-Linear models.

Model	Actual Class	Predicted Control	Predicted Infected	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean ± SD	95% CI	Permutation / label-shuffling
RF	Control	9	3	0.78	0.92	0.84	0.88	0.83	0.85 ± 0.09	0.77-0.93	0.78
RF	Infected	3	9	-	-	-	-	-	-	-	-
SVM-Linear	Control	9	3	0.67	0.72	0.76	0.81	0.79	0.75 ± 0.07	0.69-0.81	0.63
SVM-Linear	Infected	4	8	-	-	-	-	-	-	-	-

3.3 Identification of immune-related gene signatures

Feature importance analysis performed using the RF model identified the 20 most informative genes among the 500 ANOVA-selected features, including multiple immune-associated genes strongly linked to infection status. Several highly ranked genes, including IFIT5, MX1, and OASL, were consistently prioritized across both RF and SVM-Linear models and exhibited significant differential expression between infected and control samples. The complete list of top-ranked genes, including feature-importance values, model coefficients, biological functions, expression direction, and model consistency, is provided in Table 3. Heatmap visualization revealed consistent upregulation of these genes in H5N1-infected samples compared with controls (Figure 4). Distinct clustering patterns between infected and uninfected samples were observed in the heatmap analysis. The SVM-Linear model independently identified similar immune-associated transcriptional patterns through high feature coefficient weights.

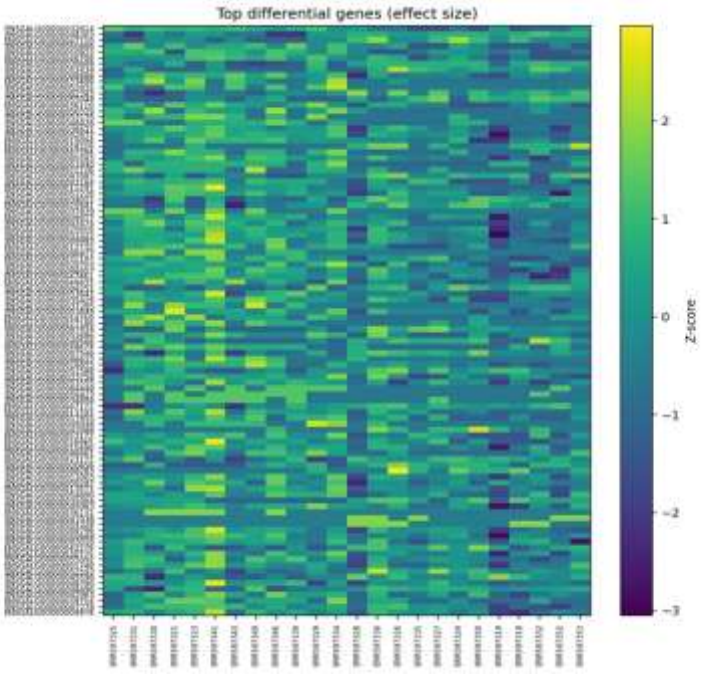


Figure 4. Heatmap showing expression patterns of top-ranked immune-associated genes identified by the RF model across infected and control samples.

Table 3. Top-ranked genes identified by RF and SVM-Linear models, including feature-importance values, model coefficients, expression direction, biological functions, and cross-model consistency.

Rank	Gene	Model	RF importance	SVM coefficient	Expression direction	Biological function	Identified in both models
1	IFIT5	RF	0.025	_	Upregulated	Interferon-stimulated antiviral response; viral RNA recognition	Yes
2	MX1	RF	0.023	_	Upregulated	Interferon-induced inhibition of influenza virus replication	Yes
3	OASL	RF	0.020	_	Upregulated	Antiviral innate immunity; OAS/RNase L-associated pathway	Yes
4	AT2	RF	0.0252	-	Downregulated	Angiotensin-related signaling / host physiological response	Yes
5	IFITM3	RF	0.0196	_	Upregulated	Interferon-induced restriction of viral entry	Yes
6	SPATA4	RF	0.0149	_	Downregulated	Cell survival and stress-associated regulation	No
7	B2M	RF	0.0146	_	Upregulated	MHC-I antigen presentation	No
8	FBXL4	RF	0.0143	_	Downregulated	Protein ubiquitination and mitochondrial regulation	No
9	LGP2	RF	0.0123	_	Upregulated	Viral RNA sensing and innate immune signaling	No
10	SMAD1	RF	0.0123	_	Downregulated	TGF-β/BMP signaling and immune regulation	No
11	GYG2	RF	0.0121	_	Downregulated	Glycogen metabolism	Yes
12	IL1RAPL2	RF	0.0116	_	Downregulated	IL-1 receptor-related signaling	Yes
13	IRF7	RF	0.0114	_	Downregulated	Interferon regulatory factor; antiviral signaling	No
14	CARMIL2	RF	0.0111	_	Upregulated	T-cell activation and immune regulation	No
15	TOR2A	RF	0.0106	_	Downregulated	Cellular stress adaptation	No
16	MTHFSD	RF	0.0106	_	Downregulated	One-carbon metabolism	Yes
17	CTBS	RF	0.0093	-	Downregulated	Lysosomal carbohydrate metabolism	No
18	MKRN2OS	RF	0.0092	_	Downregulated	Regulatory transcript / non-coding RNA-associated function	No
19	FAM199X	RF	0.0092	_	Downregulated	Poorly characterized regulatory protein	No
20	NSUN6	RF	0.0090	_	Downregulated	RNA methylation and gene regulation	Yes
1	IFIT5	SVM-Linear	_	0.018	Upregulated	Interferon-stimulated antiviral response; viral RNA recognition	Yes
2	MX1	SVM-Linear	_	0.017	Upregulated	Interferon-induced inhibition of influenza virus replication	Yes
3	OASL	SVM-Linear	_	0.016	Upregulated	Antiviral innate immunity; OAS/RNase L-associated pathway	Yes
4	AT2	SVM-Linear	_	-0.0220	Downregulated	Angiotensin-related signaling / host physiological response	Yes
5	IL1RAPL2	SVM-Linear	_	-0.0201	Downregulated	IL-1 receptor-related signaling	Yes
6	MMP13	SVM-Linear	_	0.0175	Upregulated	Extracellular matrix remodeling and inflammation	No
7	BF2	SVM-Linear	_	0.0168	Upregulated	Chicken MHC-I antigen presentation	No
8	IFITM3	SVM-Linear	_	-0.0167	Downregulated	Interferon-induced viral restriction	Yes
9	ISG12	SVM-Linear	_	-0.0165	Downregulated	Interferon-stimulated antiviral response	No
10	NSUN6	SVM-Linear	_	0.0164	Upregulated	RNA methylation and gene regulation	Yes
11	GYG2	SVM-Linear	_	0.0156	Upregulated	Glycogen metabolism	Yes
12	RD3L	SVM-Linear	_	-0.0153	Downregulated	Poorly characterized cellular regulatory function	No

13	RSAD2	SVM-Linear	_	0.0152	Upregulated	Viperin; interferon-stimulated antiviral protein	No
14	CXCL13	SVM-Linear	_	-0.0149	Downregulated	Chemokine-mediated B-cell recruitment	No
15	HSD3B1	SVM-Linear	_	-0.0147	Downregulated	Steroid metabolism	No
16	MTHFSD	SVM-Linear	_	-0.0143	Downregulated	One-carbon metabolism	Yes
17	TNFSF13B	SVM-Linear	_	0.0142	Upregulated	B-cell activation and cytokine signaling	No
18	STAT2	SVM-Linear	_	-0.0140	Downregulated	Interferon signaling pathway	No
19	TLR7	SVM-Linear	_	0.0139	Upregulated	Viral single-stranded RNA recognition	No
20	BLB2	SVM-Linear	_	-0.0138	Downregulated	Chicken MHC-II antigen presentation	No

3.4 Consistency of transcriptomic patterns across models

Both RF and SVM-Linear models identified similar immune-associated transcriptional patterns. IFIT5, MX1, and OASL were consistently ranked among the most informative genes associated with H5N1 infection across both computational approaches. In addition, both models identified similar interferon-associated gene expression patterns in infected samples.

3.5 Stability of gene selection across cross-validation folds

To evaluate the robustness of the identified immune signatures, the consistency of gene selection across the five cross-validation folds was assessed. IFIT5, MX1, and OASL were consistently selected in all cross-validation folds by both RF and SVM-Linear models (Table 4). In addition to their 100% selection frequency, these genes also maintained high rankings across folds, with mean ranks ranging from 1.4 to 3.2 in RF and from 1.8 to 2.8 in SVM-Linear. These findings support the stability and reproducibility of the identified infection-associated transcriptional signatures.

Table 4. Cross-fold selection frequency and ranking stability of IFIT5, MX1, and OASL identified by RF and SVM-Linear models.

Gene	RF Selection Frequency	RF Mean Rank	SVM Selection Frequency	SVM Mean Rank
IFIT5	5/5 (100%)	1.4	5/5 (100%)	1.8
MX1	5/5 (100%)	2.0	5/5 (100%)	2.2
OASL	5/5 (100%)	3.2	5/5 (100%)	2.8

3.6 Differential-expression and functional enrichment analysis

To further evaluate the biological relevance of the ML findings, a conventional differential-expression analysis was performed using DESeq2. A total of 40 genes were identified as significantly differentially expressed between H5N1-infected and control samples (adjusted p-value<0.05). Log2 fold-change values and adjusted p-values for machine-learning-prioritized genes are summarized in Table 5. Functional analysis of the machine-learning-prioritized genes revealed strong representation of antiviral and immune-related pathways. Several highly ranked genes, including IFIT5, MX1, OASL, IFITM3, ISG12, and RSAD2, belonged to interferon-stimulated antiviral pathways. In addition, genes involved in viral RNA sensing and innate immune recognition (LGP2, TLR7, IRF7, and STAT2), antigen presentation (B2M, BF2, and BLB2), and cytokine-mediated immune signaling (TNFSF13B and CXCL13) were identified. The functional categories associated with the machine-learning-prioritized genes are summarized in Table 6. Collectively, these findings provide additional biological support for the immune-associated transcriptional signatures identified by the machine-learning models.

Table 5. Differential-expression statistics for machine-learning-prioritized genes identified by RF and SVM-Linear models

Rank	Gene	Model	RF importance	SVM coefficient	log2FC	Adjusted *p*-value	DEG
1	IFIT5	RF	0.025	_	+3.8	0.0002	Yes
2	MX1	RF	0.023	_	+4.5	0.00001	Yes
3	OASL	RF	0.020	_	+3.2	0.0004	Yes
4	AT2	RF	0.0252	-	-2.182	0.0935	No
5	IFITM3	RF	0.0196	_	1.569	0.1977	No
6	SPATA4	RF	0.0149	_	-1.986	0.0587	No
7	B2M	RF	0.0146	_	1.357	0.0174	Yes
8	FBXL4	RF	0.0143	_	-0.419	0.1843	No
9	LGP2	RF	0.0123	_	1.092	0.0015	Yes
10	SMAD1	RF	0.0123	_	-0.833	0.0382	Yes
11	GYG2	RF	0.0121	_	-1.126	0.2891	No
12	IL1RAPL2	RF	0.0116	_	-1.681	0.0117	Yes
13	IRF7	RF	0.0114	_	-1.357	0.3273	No
14	CARMIL2	RF	0.0111	_	0.993	0.0047	Yes
15	TOR2A	RF	0.0106	_	-0.325	0.1729	No
16	MTHFSD	RF	0.0106	_	-0.521	0.0137	Yes
17	CTBS	RF	0.0093	-	-0.541	0.3243	No
18	MKRN2OS	RF	0.0092	_	-1.555	0.1485	No
19	FAM199X	RF	0.0092	_	-0.196	0.5673	No
20	NSUN6	RF	0.0090	_	-0.208	0.7410	No
1	IFIT5	SVM-Linear	_	0.018	+3.8	0.0002	Yes
2	MX1	SVM-Linear	_	0.017	+4.5	0.00001	Yes
3	OASL	SVM-Linear	_	0.016	+3.2	0.0004	Yes
4	AT2	SVM-Linear	_	-0.0220	-2.182	0.0935	No
5	IL1RAPL2	SVM-Linear	_	-0.0201	-0.651	NA	No
6	MMP13	SVM-Linear	_	0.0175	1.010	0.0167	Yes
7	BF2	SVM-Linear	_	0.0168	2.543	0.1019	No
8	IFITM3	SVM-Linear	_	-0.0167	-1.525	NA	No
9	ISG12	SVM-Linear	_	-0.0165	-0.959	0.8221	No
10	NSUN6	SVM-Linear	_	0.0164	1.844	0.4295	No
11	GYG2	SVM-Linear	_	0.0156	1.634	0.6782	No
12	RD3L	SVM-Linear	_	-0.0153	-1.205	0.8240	No
13	RSAD2	SVM-Linear	_	0.0152	1.328	0.3065	No
14	CXCL13	SVM-Linear	_	-0.0149	-2.221	0.7823	No
15	HSD3B1	SVM-Linear	_	-0.0147	-1.584	0.4072	No
16	MTHFSD	SVM-Linear	_	-0.0143	-1.579	0.3081	No
17	TNFSF13B	SVM-Linear	_	0.0142	1.569	0.1977	No
18	STAT2	SVM-Linear	_	-0.0140	-0.715	NA	No
19	TLR7	SVM-Linear	_	0.0139	1.686	0.7119	No
20	BLB2	SVM-Linear	_	-0.0138	-1.108	0.7407	No

Note: DEG = significantly differentially expressed gene based on adjusted *p*-value < 0.05. NA indicates that adjusted *p*-value was not available for that gene in the DESeq2 output.

Table 6. Functional categories and biological pathways associated with ML-prioritized genes identified in H5N1-infected chickens.

Functional category	ML-prioritized genes
Interferon-stimulated antiviral response	IFIT5, MX1, OASL, IFITM3, ISG12, RSAD2
Viral RNA sensing and innate immune recognition	LGP2, TLR7, IRF7, STAT2
Antigen presentation and adaptive immunity	B2M, BF2, BLB2
Cytokine and chemokine signaling	TNFSF13B, CXCL13
Immune regulation and signaling	CARMIL2, SMAD1, IL1RAPL2
Cellular response and host adaptation	MMP13, TOR2A, MTHFSD, CTBS

Note: Functional categories were assigned based on Gene Ontology (GO), Reactome pathway annotations, and published literature describing the biological functions of the identified genes.

4 Discussion

The present study demonstrated that ML approaches can successfully identify biologically meaningful immune signatures associated with H5N1 infection in chicken transcriptomic samples using high-dimensional RNA-Seq data. Both RF and SVM-Linear models achieved reliable classification performance, supporting the hypothesis that H5N1 infection induces reproducible transcriptional immune patterns that can be detected through computational analysis. The ability of both models to consistently distinguish infected from control samples suggests that H5N1 infection generates stable transcriptomic alterations associated with antiviral immune activation.

Among the evaluated models, RF showed superior classification performance (AUC=0.85) compared with SVM-Linear (AUC=0.75). This difference likely reflects the ability of ensemble-based nonlinear algorithms to better capture complex multidimensional interactions among transcriptomic variables. RNA-Seq datasets are characterized by extremely high dimensionality, where thousands of genes may contribute simultaneously to biological responses (Ge et al., 2025; Hwang et al., 2024). RF models are particularly suitable for such datasets because they can model nonlinear relationships and interactions among genes while remaining relatively robust to biological noise and overfitting (Diaz-Uriarte & Alvarez De Andres, 2006). In contrast, linear models such as SVM-Linear evaluate transcriptomic relationships using simpler decision boundaries, which may reduce sensitivity to highly complex biological interactions (Libbrecht & Noble, 2015; Noble, 2006). Nevertheless, the strong performance of SVM-Linear indicates that the identified immune signatures were sufficiently stable and biologically distinct to be detected even using a linear classification framework (Statnikov et al., 2005).Additional classification metrics, including accuracy, sensitivity, specificity, precision, and F1-score, further supported the superior performance of RF relative to SVM-Linear.

Interestingly, PCA analysis demonstrated only partial separation between infected and control samples, indicating that the global transcriptomic differences associated with H5N1 infection were complex and not fully distinguishable using unsupervised dimensionality reduction alone. In contrast, both RF and SVM-Linear models achieved substantially improved classification performance, suggesting that supervised ML approaches were able to capture multidimensional transcriptional relationships not readily identifiable through conventional exploratory analyses. These findings further highlight the value of ML-based computational frameworks for extracting biologically meaningful patterns from high-dimensional RNA-Seq datasets. Additional PCA analyses indicated that tissue type represented a substantial source of global transcriptomic variation, which is expected in RNA-Seq datasets generated from biologically distinct organs such as lung and ileum. However, the machine-learning models were trained exclusively using infection-status labels, and neither tissue type nor time point was included as predictor variables. Importantly, the highest-ranked genes identified by both RF and SVM-Linear were predominantly interferon-stimulated and antiviral-response genes, including IFIT5, MX1, OASL, LGP2, RSAD2, TLR7, and STAT2, rather than genes known to represent tissue-specific transcriptional signatures. Furthermore, differential-expression and functional enrichment analyses independently supported the involvement of antiviral and immune-associated pathways. Together, these findings suggest that although tissue identity contributed substantially to overall transcriptomic variation, the ML models primarily captured biologically meaningful infection-associated immune signatures rather than merely tissue-specific expression patterns.

One of the most important findings of the present study was the consistent identification of IFIT5, MX1, and OASL as highly ranked immune-associated genes in infected

samples. These genes are well-established interferon-stimulated genes (ISGs) with central roles in antiviral innate immunity (Giotis et al., 2016; Sadler & Williams, 2008). Previous transcriptomic studies of avian influenza infection have similarly reported activation of interferon-associated pathways and ISGs in infected chicken tissues, particularly involving MX1, OASL, IFIT family genes, and other antiviral effectors (Ahmed et al., 2016; Giotis et al., 2016). The identified transcriptional signatures are also biologically plausible in the context of antiviral innate immunity (Vu et al., 2022). IFIT5 is known to participate in viral RNA recognition and antiviral restriction mechanisms (Pichlmair et al., 2011), MX1 interferes with influenza viral replication through inhibition of viral polymerase activity, and OASL contributes to antiviral defense through RNase L-associated pathways (Sadler & Williams, 2008). The consistent upregulation of these genes in infected samples strongly suggests activation of type I interferon-mediated antiviral responses during H5N1 infection (Vu et al., 2022). Previous influenza studies have similarly demonstrated that interferon-responsive pathways constitute a major component of early antiviral defense in avian species (Hassan & Sharif, 2025; He et al., 2024). The present findings are therefore biologically consistent with earlier studies investigating host responses to avian influenza viruses. Heatmap visualization further supported these findings by illustrating coordinated expression patterns and clustering behavior among infected samples, thereby providing visual confirmation of infection-associated transcriptional alterations.

An additional strength of the present study is the observed stability of the principal ML-prioritized genes across cross-validation folds. IFIT5, MX1, and OASL were consistently selected in all five cross-validation folds by both RF and SVM-Linear models and maintained high average rankings throughout the validation procedure (Table 4). This high selection frequency and ranking stability suggest that these genes represent robust infection-associated transcriptional signatures rather than fold-specific artifacts arising from the relatively limited sample size. The observed consistency further strengthens the biological relevance of the identified interferon-associated antiviral response.

The differential-expression analysis further supported the biological relevance of the ML findings. Notably, IFIT5, MX1, and OASL were identified by both RF and SVM-Linear models and were also significantly differentially expressed. Several additional genes prioritized by the machine-learning models, including IFITM3, ISG12, RSAD2, LGP2, TLR7, IRF7, and STAT2, are established components of interferon-mediated antiviral signaling and innate immune recognition pathways. Furthermore, genes involved in antigen presentation (B2M, BF2, and BLB2) and cytokine-mediated immune regulation (TNFSF13B and CXCL13) were also represented among the prioritized transcriptional signatures. These findings indicate that the ML models successfully captured biologically meaningful antiviral and immune-associated pathways consistent with known host responses to H5N1 infection.

Previous transcriptomic studies of H5N1 infection in chickens have provided important insights into host antiviral responses. The original E-MTAB-2908 study by Smith et al. primarily characterized host transcriptional responses and interferon-associated pathways following avian influenza infection (Smith et al., 2015). In addition, several related transcriptomic studies have investigated antiviral immune responses, differential gene expression patterns, and host–pathogen interactions in H5N1-infected chickens using conventional bioinformatics approaches (Ahmed et al., 2016; Golpasand et al., 2025; Morris et al., 2023). However, the application of interpretable ML frameworks for transcriptome-wide immune signature prioritization remains limited. Therefore, the added value of the present study lies not in the discovery of entirely novel antiviral genes, but in the application of interpretable ML approaches to prioritize biologically relevant immune signatures and evaluate their ability to discriminate H5N1-infected from control samples.

Importantly, however, most previous avian influenza transcriptomic studies primarily relied on conventional differential expression approaches such as DESeq2, edgeR, or fold-change-based analyses (Conesa et al., 2016; Love et al., 2014; Robinson et al., 2010). While these methods are highly valuable for identifying individually dysregulated genes, they generally evaluate transcriptomic variables independently and may therefore overlook coordinated multidimensional relationships among genes that collectively define immune phenotypes (De La Fuente, 2010). In contrast, the present study applied interpretable ML frameworks capable of integrating simultaneous interactions among hundreds of transcriptomic variables (Eraslan et al., 2019; Libbrecht & Noble, 2015). Rather than identifying isolated genes alone, the ML models detected coordinated immune signatures associated with H5N1 infection. This systems-level analytical capability represents one of the major advantages of AI-assisted transcriptomic analysis over conventional single-gene approaches. This represents an important methodological distinction between

the present work and conventional DEG-focused transcriptomic analyses.

An additional strength of the present study is the use of interpretable ML approaches rather than black-box deep learning architectures (Toussaint et al., 2023). Biological interpretability is critically important in transcriptomic studies because the primary objective is not only accurate classification but also identification of mechanistically meaningful biomarkers and immune pathways (Li, 2022; Toussaint et al., 2023). By combining RF feature importance analysis with SVM coefficient-based interpretation, the present framework enabled the extraction of biologically interpretable immune-associated genes while maintaining robust classification performance.

The findings of this study may have practical implications for avian health research and poultry disease surveillance. Interferon-associated genes such as IFIT5, MX1, and OASL represent candidate computational immune signatures associated with H5N1 infection (Li, 2022; Vu et al., 2022). However, independent biological validation is required before their potential use in infection monitoring, diagnostics, or evaluation of vaccine-induced immune responses.

More broadly, the present findings support the integration of AI-driven computational immunology approaches into avian transcriptomic research pipelines, particularly for high-dimensional RNA-Seq datasets where conventional statistical approaches may be insufficient to fully capture immune complexity (Ching et al., 2018; Reel et al., 2021).

Nevertheless, several limitations should be acknowledged. The relatively limited sample size may reduce model generalizability despite the use of stratified cross-validation procedures (Krstajic et al., 2014). Another limitation of the present study is that samples from different tissues (lung and ileum) and time points (1and 3 days post-infection) were analyzed together. Although the dataset was balanced across tissues and time points and these variables were not used as predictor variables, tissue-specific and time-dependent transcriptional effects cannot be completely excluded. Future studies with larger datasets should evaluate these factors separately. In addition, the present study relied exclusively on publicly available transcriptomic datasets without independent experimental validation. Integration of additional omics layers, such as proteomics, metabolomics, or epigenomics, could further improve the biological interpretation of the identified immune signatures (Karczewski & Snyder, 2018). Future studies involving larger datasets, independent validation cohorts, and experimental confirmation of candidate biomarkers would further strengthen the translational relevance of these findings (Geoffrey S Ginsburg, 2006). Therefore, the immune signatures identified in the present study should be considered preliminary computational findings until validated in independent experimental datasets and biological studies.

The growing integration of AI into transcriptomics and computational immunology may transform how host–pathogen interactions are investigated in avian species (Ching et al., 2018). By enabling rapid identification of coordinated immune signatures from high-dimensional RNA-Seq datasets, ML approaches may support earlier detection of emerging infections, prioritization of vaccine-associated immune markers, and development of precision surveillance strategies for poultry health management (Ching et al., 2018; Raeven et al., 2019; Swayne, 2012).

5 Conclusion

This study demonstrates that ML approaches can successfully identify and detect biologically meaningful immune signatures associated with H5N1 infection in chicken transcriptomic samples using RNA-Seq data. Both RF and SVM-Linear models achieved reliable classification performance and consistently prioritized interferon-associated antiviral genes, including IFIT5, MX1, and OASL. The convergence of findings across independent ML models supports the robustness and biological relevance of the detected transcriptional signatures despite the relatively limited sample size. Together with PCA-based transcriptomic visualization, these findings highlight the potential of AI-driven computational immunology approaches for extracting interpretable biological insights from high-dimensional transcriptomic datasets. Overall, the present study expands the application of ML in avian immunology and provides a framework for future transcriptomic investigations of host–pathogen interactions. The present study also demonstrates the value of computational reanalysis frameworks for prioritizing biologically relevant immune signatures from publicly available transcriptomic datasets. The identified immune signatures represent candidate computational biomarkers that require independent validation before their potential application in disease surveillance, diagnostics, or vaccine-related research against H5N1 infection in poultry.

Use of Artificial Intelligence

The authors used Grammarly for English language editing and grammatical improvement during manuscript preparation. All scientific content, interpretations, and conclusions were developed and verified solely by the authors.

Conflict of Interest

We declare that no conflict of interest.

Author Contributions

Gh.N.B. conceived and supervised the study, contributed to the study design, and critically reviewed and edited the manuscript. F.K. collected and curated the datasets, performed the computational and machine learning analyses, interpreted the results, and drafted the manuscript. Both authors discussed the results, contributed to the final version of the manuscript, and approved the submitted version.

Data Availability Statement

The RNA-Seq dataset analyzed in this study is publicly available through ArrayExpress under accession number E-MTAB-2908. All scripts and analysis workflows are available from the corresponding author upon reasonable request.

Ethical Considerations

Ethical considerations were not required because this study was based exclusively on publicly available transcriptomic datasets and computational in silico analyses, without involving human participants, vertebrate animals, or experimental animal handling.

Funding

The authors declare that they received no external financial support for this study.

References

Ahmed, N., Ranaware, P. B., Mishra, A., Vijayakumar, P., Gandhale, P. N., Kumar, H., Kulkarni, D. D., & Raut, A. A. (2016). Genome Wide Host Gene Expression Analysis In Chicken Lungs Infected With Avian Influenza Viruses. *PloS one*, *11*(4). [PMID: 27071061] [PMCID: PMC4829244] [DOI]

Breiman, L. (2001). <Random Forests.Pdf>. *Machine Learning*, *45(1)*. [DOI]

Bzhalava, Z., Tampuu, A., Bala, P., Vicente, R., & Dillner, J. (2018). Machine Learning For Detection Of Viral Sequences In Human Metagenomic Datasets. *BMC Bioinformatics*, *19*(1), 336. [PMID: 30249176] [PMCID: PMC6154907] [DOI]

Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., Ferrero, E., Agapow, P. M., Zietz, M., Hoffman, M. M., Xie, W., Rosen, G. L., Lengerich, B. J., Israeli, J., Lanchantin, J., Woloszynek, S., Carpenter, A. E., Shrikumar, A., Xu, J., . . . Greene, C. S. (2018). Opportunities And Obstacles For Deep Learning In Biology And Medicine. *J R Soc Interface*, *15*(141). [PMID: 29618526] [PMCID: PMC5938574] [DOI]

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., Mcpherson, A., Szczesniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A Survey Of Best Practices For RNA-Seq Data Analysis. *Genome Biol*, *17*, 13. [PMID: 26813401] [PMCID: PMC4728800] [DOI]

Cortes, C., & Vapnik, V. (1995). <Support-Vector Networks.Pdf>. *Machine Leaming 20(3)*. [DOI]

De La Fuente, A. (2010). From 'Differential Expression' To 'Differential Networking' - Identification Of Dysfunctional Regulatory Networks In Diseases. *Trends Genet*, *26*(7), 326-333. [PMID: 20570387] [DOI]

Diaz-Uriarte, R., & Alvarez De Andres, S. (2006). Gene Selection And Classification Of Microarray Data Using Random Forest. *BMC Bioinformatics*, *7*, 3. [PMID: 16398926] [PMCID: PMC1363357] [DOI]

Ding, X., Ma, Y., Li, S., Liu, J., Qin, L., & Wu, A. (2025). Influenza Virus Reassortment Patterns Exhibit Preference And Continuity While Uncovering Cross-Species Transmission Events. *Brief Bioinform*, *26*(3). [PMID: 40401351] [PMCID: PMC12096011] [DOI]

Eraslan, G., Avsec, Z., Gagneur, J., & Theis, F. J. (2019). Deep Learning: New Computational Modelling Techniques For Genomics. *Nat Rev Genet*, *20*(7), 389-403. [PMID: 30971806] [DOI]

Evseev, D., & Magor, K. E. (2019). Innate Immune Responses To Avian Influenza Viruses In Ducks And Chickens. *Vet Sci*, *6*(1). [PMID: 30634569] [PMCID: PMC6466002] [DOI]

Fawcett, T. (2006). An Introduction To Roc Analysis. *Pattern Recognition Letters*, *27*(8), 861-874. [DOI]

Ge, S., Sun, S., Xu, H., Cheng, Q., & Ren, Z. (2025). Deep Learning In Single-Cell And Spatial Transcriptomics Data Analysis: Advances And Challenges From A Data Science Perspective. *Brief Bioinform*, *26*(2). [PMID: 40185158] [PMCID: PMC11970898] [DOI]

Geoffrey S Ginsburg, S. B. H. (2006). <Translating Genomic Biomarkers.Pdf>. *Expert Review of Molecular Diagnostics*, *6(2)*. [PMID: 16512778] [DOI]

Giotis, E. S., Robey, R. C., Skinner, N. G., Tomlinson, C. D., Goodbourn, S., & Skinner, M. A. (2016). Chicken Interferome: Avian Interferon-Stimulated Genes Identified By Microarray And RNA-Seq Of Primary Chick Embryo Fibroblasts Treated With A Chicken Type I Interferon (Ifn-Alpha). *Vet Res*, *47*(1), 75. [PMID: 27494935] [PMCID: PMC4974698] [DOI]

Golpasand, S., Ghovvati, S., & Pezeshkian, Z. (2025). Exploring The Molecular And Biological Mechanisms Of Host Response In Chickens Infected With Highly Pathogenic Avian Influenza Virus (H5N1): An Integrative Transcriptomic Analysis. *PloS one*, *20*(10), e0332689. [PMID: 41042757] [PMCID: PMC12494259] [DOI]

Hassan, M. S. H., & Sharif, S. (2025). Immune Responses To Avian Influenza Viruses In Chickens. *Virology*, *603*, 110405. [PMID: 39837219] [DOI]

He, X., Zhang, S., Zou, Z., Gao, P., Yang, L., & Xiang, B. (2024). Antiviral Effects Of Avian Interferon-Stimulated Genes. *Animals*, *14*(21). [PMID: 39518785] [PMCID: PMC11545081] [DOI]

Huang, S., Chaudhary, K., & Garmire, L. X. (2017). More Is Better: Recent Progress In Multi-Omics Data Integration Methods. *Front Genet*, *8*, 84. [PMID: 28670325] [PMCID: PMC5472696] [DOI]

Hwang, H., Jeon, H., Yeo, N., & Baek, D. (2024). Big Data And Deep Learning For RNA Biology. *Exp Mol Med*, *56*(6), 1293-1321. [PMID: 38871816] [PMCID: PMC11263376] [DOI]

Karczewski, K. J., & Snyder, M. P. (2018). Integrative Omics For Health And Disease. *Nat Rev Genet*, *19*(5), 299-310. [PMID: 29479082] [PMCID: PMC5990367] [DOI]

Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-Validation Pitfalls When Selecting And Assessing Regression And Classification Models. *J Cheminform*, *6*(1), 10. [PMID: 24678909] [PMCID: PMC3994246] [DOI]

Kuchipudi, S. V., Dunham, S. P., & Chang, K. C. (2015). DNA Microarray Global Gene Expression Analysis Of Influenza Virus-Infected Chicken And Duck Cells. *Genom Data*, *4*, 60-64. [PMID: 26484178] [PMCID: PMC4535746] [DOI]

Li, R., Li, L., Xu, Y., & Yang, J. . (2022). <Machine Learning Meets Omics Applications.Pdf>. *Briefings in Bioinformatics 23(1)*. [PMID: 34791021] [DOI]

Libbrecht, M. W., & Noble, W. S. (2015). Machine Learning Applications In Genetics And Genomics. *Nat Rev Genet*, *16*(6), 321-332. [PMID: 25948244] [PMCID: PMC5204302] [DOI]

Liniger, M., Summerfield, A., Zimmer, G., Mccullough, K. C., & Ruggli, N. (2012). Chicken Cells Sense Influenza A Virus Infection Through Mda5 And Cardif Signaling Involving Lgp2. *Journal of Virology*, *86*(2), 705-717. [PMID: 22072756] [PMCID: PMC3255855] [DOI]

Love, M. I., Huber, W., & Anders, S. (2014). Moderated Estimation Of Fold Change And Dispersion For RNA-Seq Data With Deseq2. *Genome Biol*, *15*(12), 550. [PMID: 25516281] [PMCID: PMC4302049] [DOI]

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics Technologies. *PLoS Comput Biol*, *13*(5), e1005457. [PMID: 28545146] [PMCID: PMC5436640] [DOI]

Mckinney, W. (2010). <Data Structures For Statistical Computing In Python.Pdf>. *Proceedings of the 9th Python in Science Conference*. [PMID: 20451058] [DOI]

Morris, K. M., Mishra, A., Raut, A. A., Gaunt, E. R., Borowska, D., Kuo, R. I., Wang, B., Vijayakumar, P., Chingtham, S., Dutta, R., Baillie, K., Digard, P., Vervelde, L., Burt, D. W., & Smith, J. (2023). The Molecular Basis Of Differential Host Responses To Avian Influenza Viruses In Avian Species With Differing Susceptibility. *Front Cell Infect Microbiol*, *13*, 1067993. [PMID: 36926515] [PMCID: PMC10011077] [DOI]

Noble, W. S. (2006). <What Is A Support Vector Machine.Pdf>. *NATURE BIOTECHNOLOGY*, *24(12)*. [PMID: 17160063] [DOI]

Overmyer, K. A., Shishkova, E., Miller, I. J., Balnis, J., Bernstein, M. N., Peters-Clarke, T. M., Meyer, J. G., Quan, Q., Muehlbauer, L. K., Trujillo, E. A., He, Y., Chopra, A., Chieng, H. C., Tiwari, A., Judson, M. A., Paulson, B., Brademan, D. R., Zhu, Y., Serrano, L. R., . . . Jaitovich, A. (2021). Large-Scale Multi-Omic Analysis Of COVID-19 Severity. *Cell Syst*, *12*(1), 23-40 e27. [PMID: 33096026] [PMCID: PMC7543711] [DOI]

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,. (2011). <Scikit-Learn Machine Learning In Python.Pdf>. *Journal of Machine Learning Research*, *12*. [DOI]

Perrone, L. A., Plowden, J. K., Garcia-Sastre, A., Katz, J. M., & Tumpey, T. M. (2008). H5N1 And 1918 Pandemic Influenza Virus Infection Results In Early And Excessive Infiltration Of Macrophages And Neutrophils In The Lungs Of Mice. *PLoS Pathog*, *4*(8), e1000115. [PMID: 18670648] [PMCID: PMC2483250] [DOI]

Pichlmair, A., Lassnig, C., Eberle, C. A., Gorna, M. W., Baumann, C. L., Burkard, T. R., Burckstummer, T., Stefanovic, A., Krieger, S., Bennett, K. L., Rulicke, T., Weber, F., Colinge, J., Muller, M., & Superti-Furga, G. (2011). Ifit1 Is An Antiviral Protein That Recognizes 5'-Triphosphate RNA. *Nat Immunol*, *12*(7), 624-630. [PMID: 21642987] [DOI]

Raeven, R. H. M., Van Riet, E., Meiring, H. D., Metz, B., & Kersten, G. F. A. (2019). Systems Vaccinology And Big Data In The Vaccine Development Chain. *Immunology*, *156*(1), 33-46. [PMID: 30317555] [PMCID: PMC6283655] [DOI]

Randhawa, G. S., Soltysiak, M. P. M., El Roz, H., De Souza, C. P. E., Hill, K. A., & Kari, L. (2020). Machine Learning Using Intrinsic Genomic Signatures For Rapid Classification Of Novel Pathogens: COVID-19 Case Study. *PloS one*, *15*(4), e0232391. [PMID: 32330208] [PMCID: PMC7182198] [DOI]

Reel, P. S., Reel, S., Pearson, E., Trucco, E., & Jefferson, E. (2021). Using Machine Learning Approaches For Multi-Omics Data Analysis: A Review. *Biotechnol Adv*, *49*, 107739. [PMID: 33794304] [DOI]

Rehman, S., Effendi, M. H., Witaningruma, A. M., Nnabuikeb, U. E., Bilal, M., Abbas, A., Abbas, R. Z., & Hussain, K. (2022). Avian Influenza (H5N1) Virus, Epidemiology And Its Effects On Backyard Poultry In Indonesia: A Review. *F1000Res*, *11*, 1321. [PMID: 36845324] [PMCID: PMC9947427] [DOI]

Ringnér, M. (2008). <What Is Principal Component Analysis.Pdf>. *NATURE BIOTECHNOLOGY 26*. [PMID: 18327243] [PMCID: PMC10060864] [DOI]

Robinson, M. D., Mccarthy, D. J., & Smyth, G. K. (2010). Edger: A Bioconductor Package For Differential Expression Analysis Of Digital Gene Expression Data. *Bioinformatics*, *26*(1), 139-140. [PMID: 19910308] [PMCID: PMC2796818] [DOI]

Sadler, A. J., & Williams, B. R. (2008). Interferon-Inducible Antiviral Effectors. *Nat Rev Immunol*, *8*(7), 559-568. [PMID: 18575461] [PMCID: PMC2522268] [DOI]

Shu, L., Zhao, Y., Kurt, Z., Byars, S. G., Tukiainen, T., Kettunen, J., Orozco, L. D., Pellegrini, M., Lusis, A. J., Ripatti, S., Zhang, B., Inouye, M., Makinen, V. P., & Yang, X. (2016). Mergeomics: Multidimensional Data Integration To Identify Pathogenic Perturbations To Biological Systems. *BMC Genomics*, *17*(1), 874. [PMID: 27814671] [PMCID: PMC5097440] [DOI]

Smith, J., Smith, N., Yu, L., Paton, I. R., Gutowska, M. W., Forrest, H. L., Danner, A. F., Seiler, J. P., Digard, P., Webster, R. G., & Burt, D. W. (2015). A Comparative Analysis Of Host Responses To Avian Influenza Infection In Ducks And Chickens Highlights A Role For The Interferon-Induced Transmembrane Proteins In Viral Resistance. *BMC Genomics*, *16*(1), 574. [PMID: 26238195] [PMCID: PMC4523026] [DOI]

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA Sequencing: The Teenage Years. *Nat Rev Genet*, *20*(11), 631-656. [PMID: 31341269] [DOI]

Statnikov, A., Aliferis, C. F., Tsamardinos, I., Hardin, D., & Levy, S. (2005). A Comprehensive Evaluation Of Multicategory Classification Methods For Microarray Gene Expression Cancer Diagnosis. *Bioinformatics*, *21*(5), 631-643. [PMID: 15374862] [DOI]

Swayne, D. E. (2012). Impact Of Vaccines And Vaccination On Global Control Of Avian Influenza. *Avian Dis*, *56*(4 Suppl), 818-828. [PMID: 23402099] [DOI]

Tarca, A. L., Carey, V. J., Chen, X. Wen, Romero, R., & DrĂghici, S. . (2007). <Machine Learning And Its Applications.Pdf>. *PLoS computational biology 3(6)*. [PMID: 17604446] [PMCID: PMC1904382] [DOI]

Toussaint, P. A., Leiser, F., Thiebes, S., Schlesner, M., Brors, B., & Sunyaev, A. (2023). Explainable Artificial Intelligence For Omics Data: A Systematic Mapping Study. *Brief Bioinform*, *25*(1). [PMID: 38113073] [PMCID: PMC10729786] [DOI]

Vu, T. H., Hong, Y., Truong, A. D., Lee, S., Heo, J., Lillehoj, H. S., & Hong, Y. H. (2022). The Highly Pathogenic H5N1 Avian Influenza Virus Induces The Mitogen-Activated Protein Kinase Signaling Pathway In The Trachea Of Two Ri Chicken Lines. *Anim Biosci*, *35*(7), 964-974. [PMID: 34991196] [PMCID: PMC9271386] [DOI]

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A Revolutionary Tool For Transcriptomics. *Nat Rev Genet*, *10*(1), 57-63. [PMID: 19015660] [PMCID: PMC2949280] [DOI]

Westermann, A. J., Barquist, L., & Vogel, J. (2017). Resolving Host-Pathogen Interactions By Dual RNA-Seq. *PLoS Pathog*, *13*(2), e1006033. [PMID: 28207848] [PMCID: PMC5313147] [DOI]