



Article

Exploring Pathogen Presence Prediction in Pastured Poultry Farms through Transformer-Based Models and Attention Mechanism Explainability

Athish Ram Das ¹, Nisha Pillai ² , Bindu Nanduri ¹, Michael J. Rothrock, Jr. ³ and Mahalingam Ramkumar ^{2,*}

¹ Department of Comparative Biomedical Sciences, College of Veterinary Medicine, Mississippi State University, Starkville, MS 39762, USA; ar2903@msstate.edu (A.R.D.); bnanduri@cvm.msstate.edu (B.N.)

² Department of Computer Science and Engineering, Mississippi State University, Starkville, MS 39762, USA; pillai@cse.msstate.edu

³ Egg Safety and Quality Research Unit, USDA-ARS U.S. National Poultry Research Center, Athens, GA 30605, USA; michael.rothrock@usda.gov

* Correspondence: ramkumar@cse.msstate.edu

Abstract: In this study, we explore how transformer models, which are known for their attention mechanisms, can improve pathogen prediction in pastured poultry farming. By combining farm management practices with microbiome data, our model outperforms traditional prediction methods in terms of the F1 score—an evaluation metric for model performance—thus fulfilling an essential need in predictive microbiology. Additionally, the emphasis is on making our model’s predictions explainable. We introduce a novel approach for identifying feature importance using the model’s attention matrix and the PageRank algorithm, offering insights that enhance our comprehension of established techniques such as DeepLIFT. Our results showcase the efficacy of transformer models in pathogen prediction for food safety and mark a noteworthy contribution to the progress of explainable AI within the biomedical sciences. This study sheds light on the impact of effective farm management practices and highlights the importance of technological advancements in ensuring food safety.

Keywords: food safety; pathogen; transformer; PageRank



Citation: Ram Das, A.; Pillai, N.; Nanduri, B.; Rothrock, M.J., Jr.; Ramkumar, M. Exploring Pathogen Presence Prediction in Pastured Poultry Farms through Transformer-Based Models and Attention Mechanism Explainability. *Microorganisms* **2024**, *12*, 1274.

<https://doi.org/10.3390/microorganisms12071274>

Academic Editor: Douglas D. Rhoads

Received: 26 May 2024

Revised: 15 June 2024

Accepted: 17 June 2024

Published: 23 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The growing concern over *Salmonella*, *Listeria*, and *Campylobacter* in poultry presents a pressing public health issue, as these pathogens lead to significant food-borne illnesses globally [1–3]. Addressing the challenge of detecting and managing these bacteria in pastured poultry farms calls for innovative strategies to ensure food safety and protect public health. Our paper introduces a machine-learning-based approach to improving the detection and control of pathogens in poultry production environments. Additionally, we present a new method in the field of explainable AI [4], offering a clear insight into the complex decision-making processes of advanced transformer models.

At the heart of our research is the microbiome—the diverse community of microorganisms inhabiting specific environments such as the gut of poultry. These microbial populations are pivotal in determining the health and disease susceptibility of their hosts. Some microbiota clusters can affect pathogen levels either by maintaining a balance that inhibits pathogenic growth or by fostering conditions conducive to an increase in pathogens [5]. Understanding these complex interactions is essential for effective pathogen risk management.

Recent research in the field of food safety and microbiome has focused on predictive microbiology [6,7] and understanding the decision-making processes of machine learning models. Standard methods for assessing feature importance, such as SHAP [8] and DeepLIFT [9], are commonly used across various datasets. However, these methods often overlook interactions within the microbiota, which can significantly influence the outcomes of these models. Consequently, simple feature importance analyses may not fully

explain the underlying biological processes. By employing transformer models, which are renowned for their ability to detect complex data patterns through attention mechanisms [10], we aim to identify microbiota clusters that indicate the presence or absence of pathogens. These models are particularly adept at uncovering data patterns and dependencies [10], making them well suited for analyzing microbiome data relationships that impact pathogen prevalence.

This study not only aims for improved pathogen prediction accuracy but also highlights the critical role of model explainability in biomedical applications. Explainability ensures the trustworthiness, transparency, and reliability of AI predictions for both poultry farmers and biologists. Our method analyzes the model's attention matrix and applies the PageRank algorithm [11] to clarify feature significance. Furthermore, we extend explainable AI techniques by applying spectral clustering for data cluster analysis and transforming attention matrices into adjacency matrices for graph-based visualizations, improving our interpretive capabilities for AI decision making.

While exploring attention weights in transformer models to demystify their "black-box" nature is not entirely new [12–14], we recognize the substantial potential of these weights. Although past studies suggest that attention weights should not be directly equated with explanations [15], the more recent literature suggests that these values can indeed be utilized for explanation [16,17]. Our approach also does not view them as direct interpretations or explanations; instead, we consider attention values as a step towards explaining a model.

By proposing the pathogen prediction method for pastured poultry farming, our paper contributes to the fields of predictive microbiology and food safety. It also paves the way for future research in explainable AI within biomedical sciences, using attention matrices to unveil the reasoning behind model decisions. The broader implications of our findings offer crucial insights into farm management practices that can drive the establishment of an optimal host microbiome that does not support food-borne pathogens and underscores the importance of novel analytical approaches in promoting food safety.

In summary, the primary objective of our study is to develop machine learning models with enhanced explainability for food safety. This research describes highly accurate machine learning models utilizing transformer architectures and introduces a novel model explanation technique that complements existing interpretation methods and shows significant promise for future applications.

2. Dataset

In the introductory section, it was highlighted that our research methodology was tested by combining two distinct datasets: the Farm Management Practices dataset and the Microbiome dataset, both of which were sourced from pastured poultry farms. These datasets were acquired from the foundational work of Hwang et al. and Rothrock et al. [18,19].

2.1. Farm Management Practices Dataset

This dataset is based on a longitudinal study conducted from March 2014 to November 2017, which covered 42 broiler flocks across 11 pastured poultry farms in the southeastern United States with flock sizes varying between 25 and 1500 birds. The data encompassed both pre-harvest (Feces and Soil) and post-harvest (Ceca, WCR-F, and WCR-P) samples. Documented by Xinran Xu et al. in 2021 [20], the study collected data on poultry farms using movable pens, which were shifted to new pastures daily. The configuration, number, and application of temporary fencing around the pens varied across the farms.

The data collection encompassed 40 key farm practice variables across the lifecycle of a flock, totaling 160 variables, including metadata. The dataset includes variables for the presence of *Salmonella*, *Listeria*, and *Campylobacter*, which are used as the target variables for our study.

2.2. Microbiome Dataset

To generate the Microbiome dataset, 16S rRNA gene high-throughput Illumina sequencing was applied to analyze temporal pre-harvest samples, including feces and soil, from the same 41 pastured poultry flocks. This sequencing facilitated the determination of the relative abundance of operational taxonomic units (OTUs). The unique genera identified within the OTUs served as machine learning predictors for assessing the prevalence of food-borne pathogens—*Salmonella*, *Campylobacter*, and *Listeria*—at various growth stages of poultry, categorized as START (2–4 weeks old), MID (5–7 weeks old), and END (8–11 weeks old).

Microbiome Analysis

DNA Extraction: DNA was extracted from samples according to a semi-automated hybrid DNA extraction protocol [21]. An enzymatic method based on the QIAamp DNA Stool Mini Kit (QIAGEN, Valencia, CA, USA) was combined with a mechanical method with the FastDNA Spin Kit for Feces (MP Biomedicals, Solon, OH, USA). A QIAcube Robotic Workstation was used to purify DNA using the DNA Stool-Human Stool-Pathogen Detection Protocol. Using a Take3 plate and the Synergy H4 multimode plate reader (BioTek, Winooski, VT, USA), the DNA concentration of each sample was determined spectrophotometrically after purification.

Illumina MiSeq Library Analysis: A dataset from the Earth Microbiome Project Laboratory at the U.S. Department of Energy, Argonne National Laboratory (Argonne, IL, USA) was used for library construction and sequencing. The hypervariable V4 domain of the bacterial 16S rRNA gene was amplified using the F515 (5'-CACGGTCCGKCGGCCATT-3') and R806 (5'-GGACTACHVGGGTWTCT AAT-3') primer set, with each primer containing Illumina adapter regions (Illumina, Inc., San Diego, CA, USA) and the reverse primer containing the Golay barcodes to facilitate multiplexing [22]. The Illumina MiSeq platform was used to obtain raw reads. The QIIME v1.9.1 pipeline (Quantitative Insights Into Microbial Ecology) generated and processed 3,297,242 raw sequence reads [23]. R1 reads were filtered for quality, and libraries were split according to the Golay barcode sequences (split library *fastq.py* script, default parameters). With the *usearch* option [24] and the *pick_otus.py* script (*-m usearch*, all other parameters were set to the default), sequences were chimera checked against the gold.fa database (<http://drive5.com/uchime/gold.fa>, accessed on 20 May 2024) and clustered based on their sequence similarity (97%) into operational taxonomic units (OTUs). For each OTU, a representative sequence was selected using the *pick_rep_set.py* script (utilizing the most abundant method for picking, all other parameters were set to the default) and used for taxonomic assignment with UCLUST and the Greengenes 13_8 database [25] using *assign_taxonomy.py* (default parameters). Using PyNAST (<http://pynast.sourceforge.net>, accessed on 10 December 2014) [26], sequences were aligned (*align_seqs.py* script, default parameters) and filtered (*filter_alignment.py*). Following this, a phylogenetic tree was generated using *make_phylogeny.py* (using the default settings and FastTree).

In the context of machine learning, the dataset incorporates input features derived from two distinct sources: the Farm Management Practices dataset and the Microbiome dataset. Specifically, the Farm Management Practices dataset contributes 157 input features, while the Microbiome dataset provides 1823 input features. The model targets three outcome variables, corresponding to the prevalence of three food-borne pathogens: *Salmonella*, *Campylobacter*, and *Listeria*.

3. Method

The methodology of this study is structured into three primary stages: data pre-processing, model architecture design and training, and model decoding. While these stages are commonly found in any machine learning project, our approach necessitates further elaboration due to the specific adaptations made to accommodate the transformer model.

3.1. Data Pre-Processing

In addressing the farm management data, which comprise both categorical and continuous variables, a critical step in our data pre-processing involved encoding these variables into fixed-dimension vectors. This encoding is essential for facilitating the computation within the attention mechanism of the transformer model. Drawing insights from the work of Xin Huang [27], we observed that traditionally, attention mechanisms in the context of tabular data primarily accommodated categorical features, leaving continuous features to bypass the attention blocks.

To effectively incorporate continuous features into the attention mechanism, we developed a new encoding technique named “scalar-to-vector encoding”. The details of this method are provided below.

Given a scalar value S , a maximum value max , a minimum value min , and a target vector dimension n , we first define the bin size as follows:

$$binsize = \frac{max - min}{n - 1}$$

The bin boundaries spanning the complete range of our data are set according to this $binsize$. The scalar value S is then mapped to a corresponding n -dimensional vector V where all elements are initially set to zero. To embed S into V , we identify the immediate lower L and upper U bin boundaries that S falls between. The vector V is then updated at indices corresponding to L and U as follows:

$$V_L = \frac{U - S}{U - L}$$

$$V_U = \frac{S - L}{U - L}$$

All other elements in V remain at zero, ensuring that V represents the scalar value S with respect to its position within the specified range.

This method allows continuous variables to be transformed into a fixed-dimensional vector format, making them compatible for processing through the attention mechanism. This extends the capabilities of the transformer model to accommodate scalar continuous values as well. Our design builds upon the foundational principles outlined by Xin Huang [27], expanding its utility to include continuous data alongside categorical inputs.

3.2. Model

Utilizing the scalar-to-vector encoding mentioned in the previous subsection, we merged categorical features from the Farm Management Practices dataset and continuous features from both the Farm Management Practices dataset and the Microbiome dataset, preparing them for introduction into the transformer model, as shown in Figure 1. Adopting the transformer encoder architecture outlined in “Attention is All You Need” by Vaswani et al. (2017) [10], we applied it to predict the binary presence of pathogens.

3.3. Model Decoding

In the field of model explanation, Shapley Additive Explanations (SHAP) [8] and Deep Learning Important Features (DeepLIFT) [9] are established methods. Our motivation for creating a new model explainability approach stems from two factors. Firstly, the performance of transformers across different data types has been better, a benefit linked to their attention mechanism. This mechanism implies that analyzing the attention matrix can unlock various possibilities in Explainable AI.

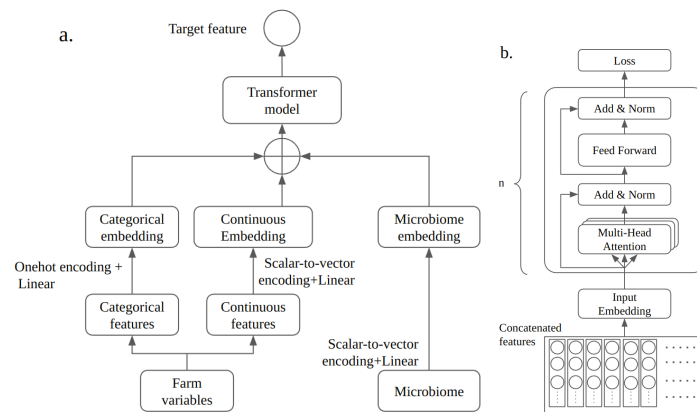


Figure 1. Illustration of the architecture of the transformer model highlighting the integration of both categorical (via one-hot encoding) and continuous (via scalar-to-vector encoding) features into the transformer’s encoder blocks. Panel (a) shows the encoding processes, and Panel (b) shows the transformer model’s encoder architecture.

Our transformer model features multiple layers of multi-head attention networks in series, with each layer’s output feeding into the subsequent layer’s input. For our purposes in explainable AI and to keep the complexity manageable, only the attention weights from the first block were used for further analysis. Thus, minor modifications in the formula for computing attention [15] allow us to derive the attention matrices for our analysis; these are represented as follows:

$$A(Q_{l1}, K_{l1}, V_{l1}) = \text{softmax}\left(\frac{Q_{l1}K_{l1}^T}{\sqrt{d_k}}\right)V_{l1}$$

Here, A represents the attention value, and Q , K , and V denote the Query, Key, and Value, respectively, with the subscript $l1$ signifying that these components are exclusively derived from the first layer. This approach is used to construct the attention matrix essential for our model decoding and is utilized for three distinct analyses.

Feature Importance with PageRank: We initially applied this approach to calculate feature importance, a common practice in decoding black-box models. PageRank is an algorithm used by Google Search to rank web pages in their search engine results [11]. It measures the importance of website pages by counting the number and quality of links to a page to determine a rough estimate of a website’s importance. Analogously, by using the PageRank algorithm, the significance of a feature can be calculated based on its attention weights.

The PageRank vector, PR , was determined by solving the following eigenvalue problem:

$$PR = \alpha APR + \frac{1 - \alpha}{N} \mathbf{1}$$

where

- PR is the vector of PageRank values for all features;
- α is the damping factor;
- A is the attention matrix, where element A_{ij} denotes the softmax attention weight of feature i to feature j ;
- N is the total number of features;
- $\mathbf{1}$ is a vector with all elements equal to 1.

Spectral Clustering to Identify the Signature of Microbiota: The attention mechanism simplifies the process of identifying feature clusters. These clusters, which are feature groups affecting the model’s output together, are more directly observable in the attention matrix. The importance of identifying such clusters is particularly relevant in our data, as

recognizing microbiota clusters is significant in microbiology. Spectral clustering utilizes the spectral features of attention weights to achieve this goal.

Spectral clustering [28] is a technique used in machine learning and data mining to identify clusters in data based on the spectrum (eigenvalues) of the similarity matrix of the data. It works by transforming the data into a lower-dimensional space in which clusters are more apparent and can be easily identified using traditional clustering techniques such as K-means [29] or HAC (Hierarchical Agglomerative Clustering) [30], the latter of which we used in our study.

Our method involved constructing a similarity matrix, S , from the data and then deriving the Laplacian matrix, L , from S . HAC was then performed in the space spanned by the eigenvectors of L corresponding to its smallest eigenvalues.

Additionally, these attention matrices can be mapped out in a graph structure to visualize the interactions among the identified OTUs in the microbiota, which is an interesting research question in microbial ecology.

4. Experiments and Results

Our investigation focused on two primary questions: whether a prediction model using a transformer architecture, as depicted in Figure 1, outperforms multi-layer perceptron (MLP) models and whether the attention matrix can offer valuable insights into explainable AI. To address the first question, we carried out extensive testing across several models using our dataset, applying a grid search to fine-tune the hyperparameters. The findings from these tests were used in evaluating the Tab-transformer model's performance relative to that of MLP models. For the second question, we extracted and analyzed the attention matrix from our model, and a detailed discussion of this analysis is presented in the following sections.

4.1. Transformer Model Evaluation

For model evaluation, we conducted a series of detailed experiments involving various combinations of epochs, learning rates, test sizes, dropout values, attention layers, and linear layers. These combinations were applied to predict pathogen presence in the following different contexts:

- Pre-harvest *Salmonella* samples;
- Pre-harvest *Listeria* samples;
- Pre-harvest *CampyCapetown* samples;
- Post-harvest *Salmonella* samples;
- Post-harvest *Listeria* samples;
- Post-harvest *CampyCapetown* samples.

We executed the experiments across the following four distinct model architectures to gauge their effectiveness:

- Multi-layer perceptron (MLP) - Farm Management Practice Variables.
- Multi-head transformers without scalar-to-vector embedding—Farm Management Practice variables.
- Multi-head transformers with scalar-to-vector embedding—Farm Management Practice variables.
- Multi-head transformers with scalar-to-vector embedding—Farm Management Practice variables and Microbiome.

This approach allowed us to compare the efficiency and accuracy of each architecture under similar experimental conditions.

The results of these experiments are shown in Table 1. The table presents the F1 scores for all the models tested on our dataset, demonstrating the gradual enhancement in performance as the models evolve. Our final architecture incorporates a multi-head transformer with scalar-to-vector encoding for continuous variables. The dataset used for this final model combines poultry management variables and microbiome data.

The results of these experiments demonstrate that the performance of the transformer models is better than that of the multi-layer perceptrons. Moreover, the model gives much better predictions of pathogen presence when the farm management data are combined with the Microbiome dataset. With these results, we can infer that the attention mechanism is working in our model, which is essentially the core of transformer models, and we were ready to further investigate the attention matrix.

Table 1. F1 scores for all the models tested on the dataset.

	Pre Harvest			Post Harvest		
	<i>Salmonella</i>	<i>Listeria</i>	<i>Campy</i>	<i>Salmonella</i>	<i>Listeria</i>	<i>Campy</i>
MLP	0.79	0.67	0.84	0.78	0.87	0.95
Multi-Head Transformer w/o scalar-to-vector embedding	0.79	0.72	0.84	0.78	0.92	0.96
Multi-Head Transformer w/ scalar-to-vector embedding	0.78	0.71	0.84	0.83	0.91	0.97
Multi-Head Transformer w/ scalar-to-vector embedding	0.86	0.79	0.86	0.89	0.92	0.97

4.2. PageRank Results and Evaluation

As described in the Methods section, we extracted the attention weights from the first multi-head attention block's self-attention heads for all test scenarios. We then applied the PageRank algorithm to allocate scores to each feature, with the aggregate of these scores equaling 1 amongst all 1823 microbiota features and around 60 farm variable features. The top PageRank-valued features were identified for validation. The top 10 microbiota features identified by the method for post-harvest *Salmonella* are shown in Figure 2 as an example. Similar tables were generated for pre-harvest and post-harvest *Salmonella*, *Listeria*, and *CampyCapetown*.

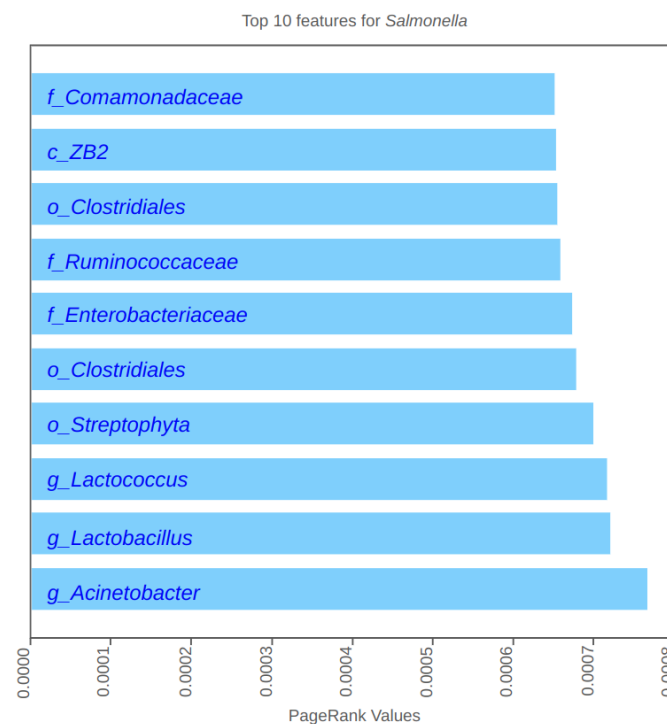


Figure 2. Top 10 most important features recognized by PageRank for post-harvest *Salmonella*.

The outcomes were verified through both qualitative and quantitative means. For the qualitative evaluation, we cross-referenced the top microbiota features identified by our approach with the existing literature to confirm their associations with probiotic or pathogenic characteristics, as discussed in detail in Section 5. For quantitative evaluation, we used DeepLIFT to rank all features in the dataset and then compared the top 100 features identified by both our method and DeepLIFT. This comparison revealed a significant concurrence between the two methods, with approximately 35% of the top features being recognized by both. The Venn diagram in Figure 3 illustrates some of these results. For a comprehensive view of the findings, please refer to the Appendix A.

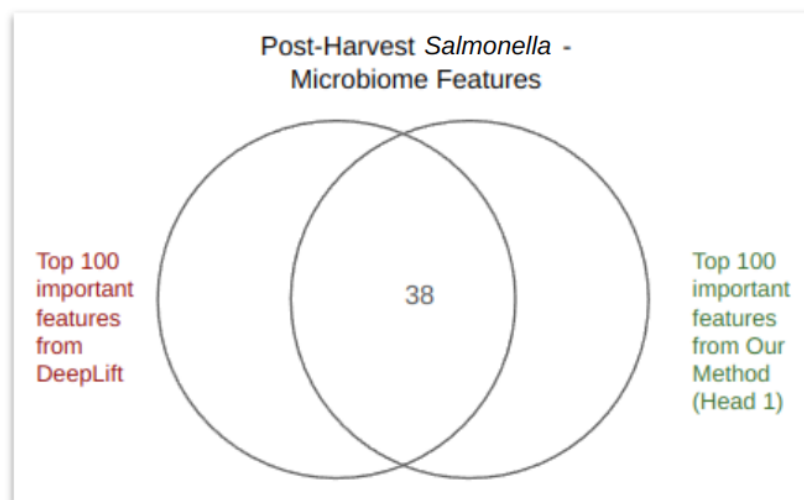


Figure 3. Comparison of the DeepLIFT and PageRank results for post-harvest *Salmonella*. The top 100 features from a total of 1824 were chosen to construct the Venn diagram, which reveals a notable level of agreement between the two methods.

4.3. Spectral Clustering Results

To determine microbiota clusters within the six experimental scenarios (Section 4.1), we utilized attention matrices. The selection of the number of clusters, a critical hyperparameter in our clustering approach, necessitated a methodical decision-making process. We opted for Hierarchical Agglomerative Clustering (HAC) over K-means due to the clarity provided by HAC's dendrograms in determining the optimal number of clusters. For each scenario, a dendrogram was generated to establish the number of clusters. Subsequently, we applied HAC with average linkage for feature classification into clusters. With this approach, we were able to identify clusters composed of microbes with similar ecological properties. An example of two clusters identified through this method is shown in Table 2. In one of the clusters, the majority of species, including *Actinobacteria*, *Acidobacteria*, *Bacteroidetes*, *Rhodoplanes*, *Bacillus*, *Myxococcales*, and *Candidatus Nitrososphaera*, are non-pathogenic and beneficial microorganisms primarily distributed in soil and water environments. These groups play crucial roles in ecosystem functioning, such as in nutrient cycling and organic matter decomposition. In contrast, the *Lactobacillales* found in the other cluster are typically associated with environments rich in carbohydrates, including dairy products, fermented foods, the gastrointestinal tract of humans and animals, and plant surfaces.

Table 2. Evaluation of two clusters identified through HAC. Microbes usually found in soil (cluster B/2) and gut (cluster A/1) are accurately identified and grouped separately.

Bacteria	Cluster
<i>g__Rummeliibacillus</i>	2
<i>g__Salinicoccus</i>	2
<i>g__Enterococcus</i>	2
<i>g__Streptococcus</i>	2
<i>f__Enterobacteriaceae;g__</i>	2
<i>g__Candidatus Nitrososphaera</i>	0
<i>o__iii1-15;f__g__</i>	0
<i>o__RB41;f__g__</i>	0
<i>f__Ellin6075;g__</i>	0
<i>f__Gaiellaceae;g__</i>	0
<i>o__Solirubrobacterales;f__g__</i>	0
<i>f__Solirubrobacteraceae;g__</i>	0
<i>f__Cytophagaceae;g__</i>	0
<i>o__Sphingobacteriales;f__g__</i>	0
<i>f__Chitinophagaceae;g__</i>	0
<i>g__Bacillus</i>	0
<i>g__Rhodoplanes</i>	0
<i>o__Myxococcales;f__g__</i>	0
<i>g__DA101</i>	0

5. Discussion

This research aims to make powerful machine learning models more interpretable and transparent when analyzing complex biological data. By combining techniques to explain these models, using effective transformer neural networks, and applying algorithms such as PageRank, we can gain deeper insights into fundamental biological processes. This approach has wide applications to other animal production systems beyond the pastured poultry production system used in this study. It can also be applied to drug discovery, analyzing evolutionary relationships between species, mining agricultural/biomedical text data, understanding gene regulatory networks, and predicting protein structures. Making these advanced models more interpretable will help advance our understanding of biology and lead to new applications that can improve animal and human health.

This section presents a discussion of our findings and related research, focusing on microbiota features that have been identified as significantly associated with the presence of food-borne pathogens. *Salmonella* and *Campylobacter* are among the most prevalent pathogens associated with poultry and are leading causes of bacterial food-borne illness [31]. Taxonomically, *Salmonella* belongs to the phylum *Proteobacteria*, while *Campylobacter* belongs to the phylum *Epsilonbacteraeota* [32]. *Listeria*, another food-borne pathogen, is a concern not only in poultry but also in a wide range of other foods, such as dairy and ready-to-eat products, and it belongs to the phylum *Firmicutes* [33]. However, the presence and significance of *Listeria* in poultry production, particularly in pastured systems, is more associated with the environment and processing facilities than with the birds themselves.

5.1. Role of Firmicutes

Our research finds *Firmicutes* to be one of the major influencing factors for the prevalence of *Salmonella*, *Listeria*, and *Campylobacter* in both pre-harvest (feces, soil) samples and post-harvest (post-processing) samples. While *Firmicutes* is a large phylum of bacteria that includes many non-pathogenic and beneficial organisms, the environmental dynamics of pastured poultry systems can affect the prevalence of various pathogens, including those not classified under *Firmicutes* [34]. Some beneficial *Firmicutes* in the gut flora of pastured

poultry may help in competitively excluding pathogenic bacteria by competing for nutrients and attachment sites in the gastrointestinal tract [35]. Studies show that inoculation with *S. Enteritidis* [36] resulted in significant positive correlations with *Firmicutes*, notably affecting the relative abundance of 18 genera.

We found that the family *Bacillaceae*, belonging to the phylum *Firmicutes* [32], influenced the prevalence of *Salmonella*, *Listeria*, and *Campylobacter* in our experiments. The ability of *Bacillus* strains to produce a wide array of antimicrobial peptides (AMPs) and bacteriocins is crucial in their antagonistic effects against enteropathogenic bacteria in the gastrointestinal tract [37]. Specifically, the growth of *Listeria* is inhibited in contaminated environments or products [38]. *Bacillus subtilis* PS-216 showed effective inhibition against *Campylobacter jejuni* under microaerobic conditions, demonstrating its potential as a probiotic that could integrate into the chicken intestinal microbiome and combat *campylobacteriosis* [39]. *Bacillus* strains can produce extracellular polysaccharides, vitamins, and exoenzymes that support the growth of beneficial microbiota, contributing to a healthier gut environment [40,41] and potentially reducing the colonization of pathogens by strengthening the birds' natural defenses against infections and possibly reducing pathogen shedding.

Our results show that *Rummeliibacillus*, a Gram-positive rod-shaped bacterium, is associated with *Listeria* and *Campylobacter*. Factors such as soil composition and microbial diversity can either facilitate or limit the survival of *Listeria* [42], indicating that microbial competition, potentially including competition from *Rummeliibacillus*, could influence the prevalence of *Listeria*. Lactic acid bacteria (LAB) [43], including *Lactococcus* [44] and *Lactobacillus* species, are known for their role in producing fermented foods and for their ability to inhibit the growth of pathogenic bacteria through the production of lactic acid, bacteriocins, and AMPs. Acidification of the environment due to lactic acid generation by LAB inhibits the growth of [45,46] *Listeria* and *Salmonella*. *Lactobacillus* species have been extensively studied for their probiotic properties in poultry [47], demonstrating benefits such as reduced *Salmonella* contamination (<https://today.uconn.edu/2021/06/probiotic-intervention-to-prevent-salmonella-infection-in-poultry/> accessed on 8 April 2024), improved growth performance, immune enhancement, gut microbe sustainability, and contributions to health. *Lactobacillus* cultures or bacteriocins could be used in rinses or coatings for poultry meat post-processing to reduce surface contamination by *Salmonella* [48].

Our results indicate the influence of *Lysinibacillus* species on the presence of *Salmonella* and that of the *Planococcaceae* family on the presence of *Listeria* in the pre-harvest phase, and this suggests a multifaceted approach involving the management of animal waste, the monitoring and treatment of irrigation water, and practices to reduce contamination in food production environments. The role of soil-dwelling or fecal bacteria, such as *Lysinibacillus* [49], in influencing these processes, directly or indirectly, through effects on microbial communities remains a critical area for further research. Another influential genus in our findings, *Solibacillus*, a Gram-positive, rod-shaped, spore-forming bacteria, could potentially compete with *Salmonella* for nutrients in the soil, alter soil microbial community composition, and limit the latter's ability to proliferate.

Anoxybacillus is a genus of thermophilic [50], facultatively anaerobic bacteria within the *Firmicutes* phylum that is known for its ability to thrive at high temperatures and for its presence in diverse environments, including hot springs and dairy products. Detecting *Anoxybacillus* in post-harvest environments indicates that higher temperatures or specific nutrient availability could impact the survival or proliferation of pathogens such as *Campylobacter*.

The class *Clostridia*, part of the *Firmicutes* phylum, includes significant food-borne pathogens that impact poultry and can pose risks to human health [51]. Our results indicate that *Clostridia* plays a critical role in the prevalence of *Salmonella* and *Campylobacter* in soil and fecal samples. Conversely, the presence of *Salmonella* and *Listeria* was associated *Clostridia* species in post-processed meat samples. By contributing to the fermentation

process and production of short-chain fatty acids (SCFAs) in the gut [52], The *Ruminococcaceae* family, part of the *Clostridia* class, might help create an intestinal environment that is less favorable for *Salmonella* and *Campylobacter* proliferation. The association between the *Syntrophomonas* genus and *Listeria* contamination in our results suggests a potential role for this group of anaerobic, syntrophic bacteria in modulating the prevalence of *Listeria* in the post-processing environment, possibly due to the syntrophic degradation of butyrate in environmental and industrial processes limiting organic waste that can support pathogens such as *Listeria* [53].

5.2. Role of Proteobacteria

The phylum *Proteobacteria* represents a vast and diverse group of Gram-negative bacteria that are classified into various classes. *Acinetobacter* can be found in soil and animal (including human) feces, though they are not a predominant component of the gut flora [54], so they could indicate exposure to the environment or the consumption of contaminated food or water [55–57]. In soil, these bacteria are key players in breaking down organic substances and nutrient recycling [58].

Our previous studies have underscored the potential role of other microbial species, such as *Acinetobacter*, in the ecology of these pathogens. Specifically, our analysis indicates that *Acinetobacter* played a critical role in pre-harvest (fecal and soil) samples where *Salmonella* was detected. Detection of *Salmonella* and *Acinetobacter* from poultry operations in Washington state was reported [59]. Understanding the nature of this relationship and whether it is synergistic or antagonistic could inform the development of targeted interventions and more effective pathogen control strategies on farms.

5.3. Clustering to Uncover the Microbial Community Structure

ClusterA

Rummeliibacillus is a part of the *Firmicutes* phylum, and these Gram-positive bacteria are known to form endospores, which allow them to survive in harsh environmental conditions. *Streptococcus* is another member of the *Firmicutes* phylum; these are Gram-positive cocci known for their role in both health (as part of the normal microbiota) and disease (causing various infections). *Enterococcus*, *Streptococcus*, and members of the *Enterobacteriaceae* family are primarily associated with the gastrointestinal tract of animals and humans [60], playing roles that range from benign colonization to causing serious infections. *Salinicoccus* is a genus of Gram-positive, halophilic bacteria that belong to the family *Staphylococcaceae* [61]. These bacteria are typically found in environments with high salt concentrations [62], such as salt lakes, saline soils, and salted food products, and they have potential applications in biotechnology, including the biodegradation of pollutants in saline conditions and the production of enzymes and other bioactive compounds. *Enterococcus*, including species such as *Enterococcus faecalis*, is a genus of bacteria that are part of the natural microbiota of the human gastrointestinal tract [63] but can also be found in soil, water, food, and decaying vegetation [64]. In soil, *Enterococcus* species may be introduced through the application of animal manure as fertilizer, contributing to the microbial diversity of agricultural soils.

ClusterB

All of the groups in this cluster are ubiquitous in the environment and are found in a wide range of habitats from soil and water to extreme environments such as hot springs (*Crenarchaeota*) and acidic mines (*Acidobacteria*). This distribution underlines their adaptability and the vast diversity of metabolic strategies that they have evolved to exploit different ecological niches. They play crucial roles in their respective ecosystems and are involved in nutrient cycling, decomposing organic matter, and, in some cases, forming symbiotic relationships with plants [65] (e.g., certain *Alphaproteobacteria*, such as *Rhizobia*) or animals. Their metabolic diversity allows them to perform a variety of biochemical processes critical to Earth's biogeochemical cycles, such as carbon and nitrogen cycling.

6. Conclusions

In this study, we set out to investigate two main questions: firstly, whether incorporating a transformer model into our unique framework would significantly enhance our ability to predict pathogen presence in poultry production environments; secondly, whether we could introduce a novel method for determining feature importance using the attention matrix and PageRank. Although our final model showed modest improvements in predictive performance over previous models applied to our dataset, this improvement was attributed more to the inclusion of microbiome data than to the transformer model itself. However, the primary goal of implementing the transformer model was to facilitate our second objective, which indeed yielded promising results. The method of computing feature importance through the attention matrix and PageRank aligned well with DeepLIFT's findings and was validated as biologically relevant. Additionally, our study introduced an effective method for identifying microbiota signatures using the same attention matrices, yielding meaningful outcomes. Consequently, we view our contribution as a positive step forward in the field of explainable AI, and we anticipate that it will inspire further research in this direction.

Author Contributions: Conceptualization, A.R.D. and N.P.; methodology, A.R.D. and M.J.R.J.; software, A.R.D.; validation, A.R.D.; formal analysis, A.R.D.; investigation, A.R.D. and M.J.R.J.; resources, M.J.R.J. and B.N.; data curation, M.J.R.J. and A.R.D.; writing—original draft preparation, A.R.D. and N.P.; writing—review and editing, A.R.D., B.N., N.P., M.J.R.J. and M.R.; visualization, A.R.D.; supervision, M.J.R.J. and B.N.; project administration, B.N. and M.J.R.J.; funding acquisition, B.N. and M.J.R.J. All authors have read and agreed to the published version of the manuscript.

Funding: The dataset used in this study was provided by the Agricultural Research Service USDA CRIS Project “Reduction of Invasive Salmonella enterica in Poultry through Genomics Phenomics and Field Investigations of Small Multi-Species Farm Environments” #6040-320000-011-00-D. This work was supported by the USDA-ARS NACA Project “Developing Detection and Modeling Tools for the Geospatial and Environmental Epidemiology of Animal Disease” #58-6064-3-017.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from the Agricultural Research Service, USDA, and are available from Dr. Rothrock with the permission of the USDA.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Appendix A.1. Transformer Model Performance

The performance of the prediction models tested on our dataset through the evolution of our models is shown in Figure A1.

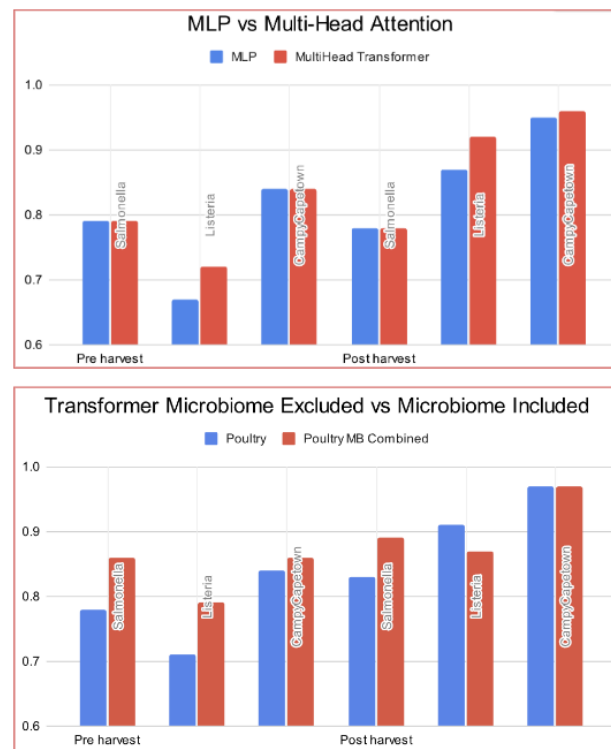


Figure A1. Comparison of the average F1 scores on the test set for predictions made by the different models.

Appendix A.2. PageRank Results

The top 10 features identified with the PageRank method for different scenarios are shown in Figures A2 and A3. Figures A4 and A5 show the agreement in results for both DeepLIFT and our method across different scenarios for the microbiota features. A similar agreement was also found in the variables of the Farm Management Practices dataset, and this is shown in Figure A6.

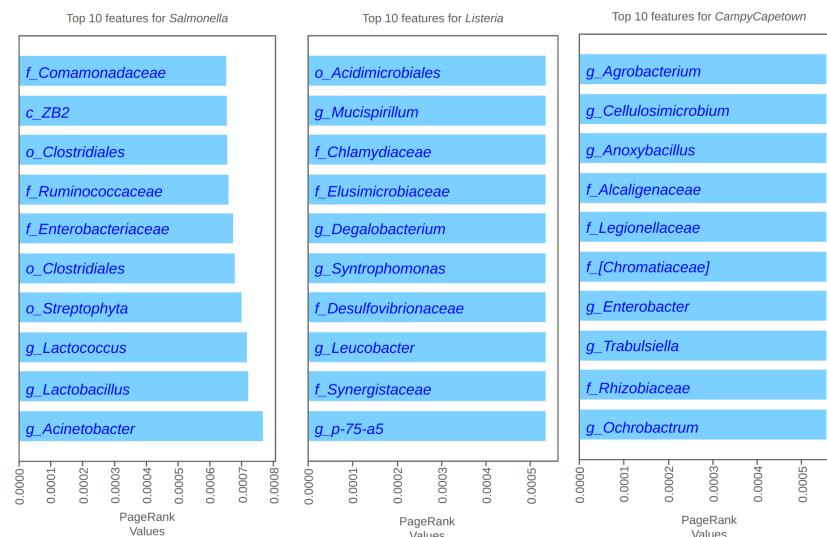


Figure A2. Top 10 features identified with PageRank for post-harvest *Salmonella*, *Listeria*, and *CampyCapetown*.

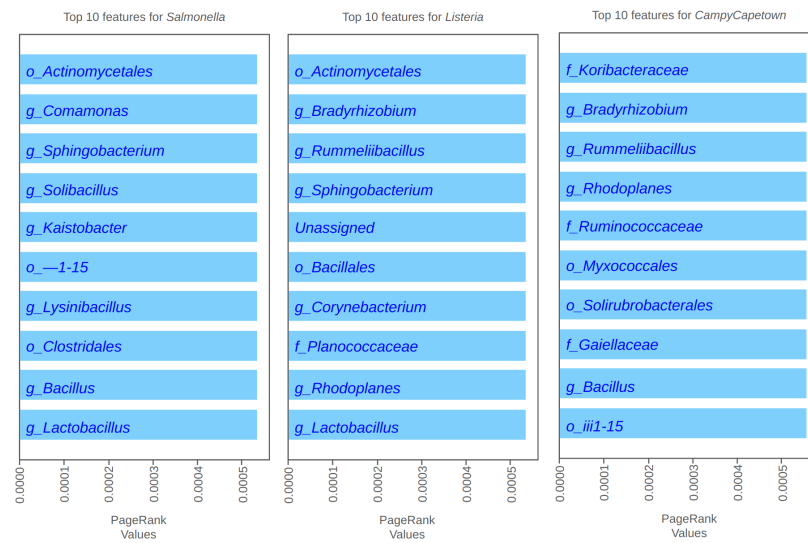


Figure A3. Top 10 features identified with PageRank for pre-harvest *Salmonella*, *Listeria*, and *Campy-Capetown*.

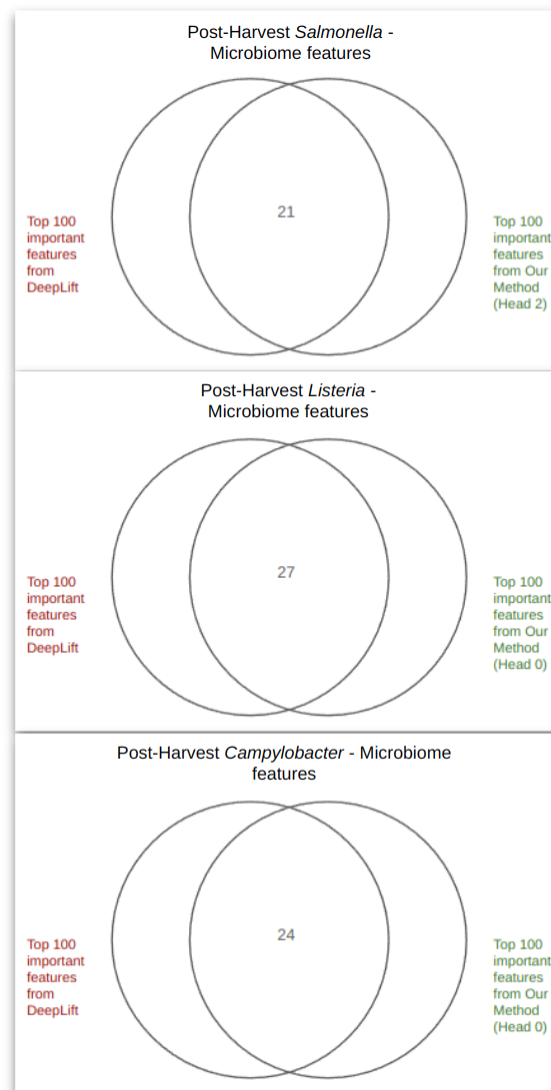


Figure A4. Comparison of DeepLIFT and PageRank. The feature importance identification across different post-harvest scenarios shows a notable agreement.

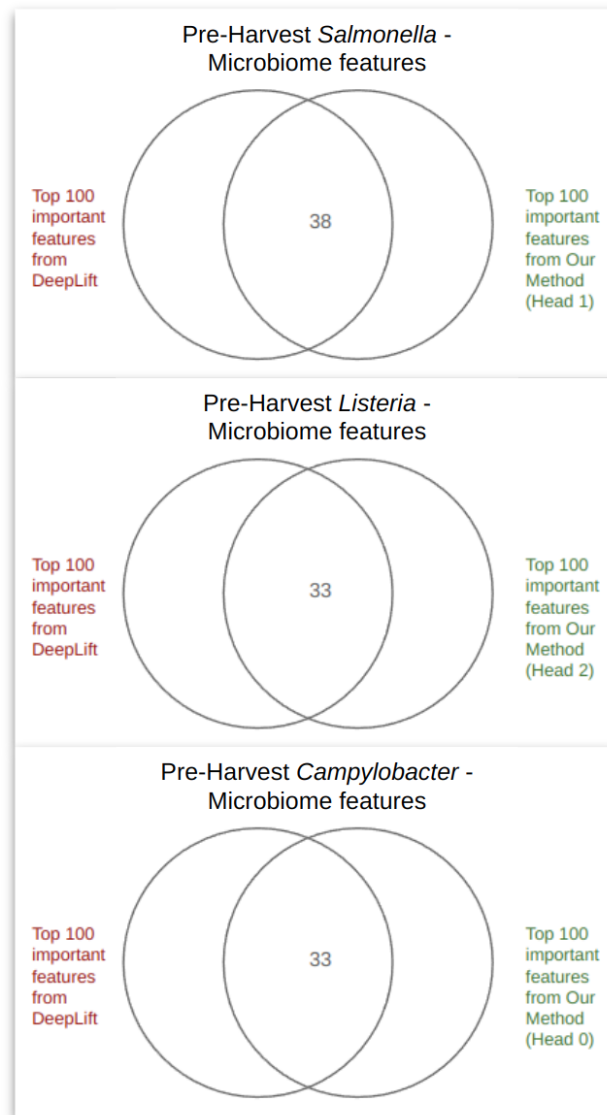


Figure A5. Comparison of DeepLIFT and PageRank. The feature importance identification across different pre-harvest scenarios shows a notable agreement.

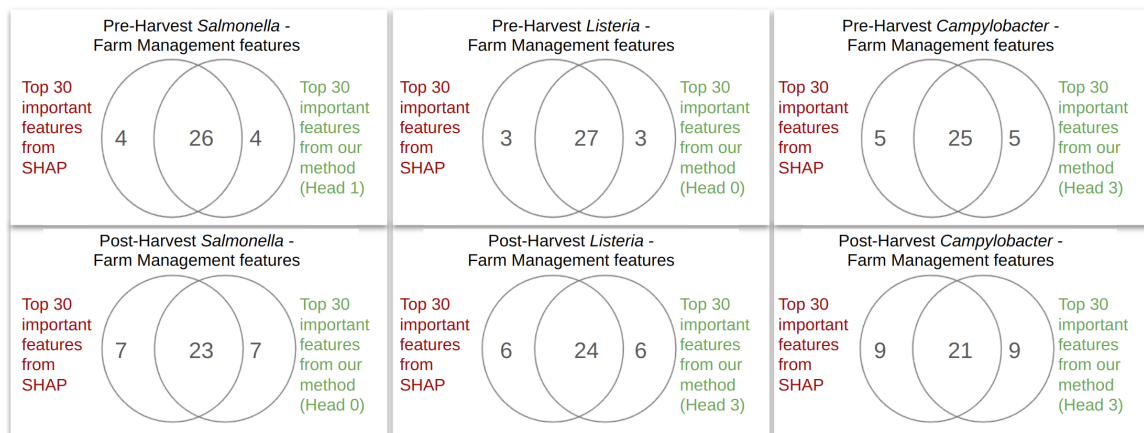


Figure A6. Comparison of DeepLIFT and PageRank. The feature importance identification across all scenarios for farm management features shows a notable agreement.

References

1. Rothrock, M.J.; Davis, M.L.; Locatelli, A.; Bodie, A.; McIntosh, T.G.; Donaldson, J.R.; Ricke, S.C. Listeria Occurrence in Poultry Flocks: Detection and Potential Implications. *Front. Vet. Sci.* **2017**, *4*, 125. [[CrossRef](#)]
2. DuPont, H.L. The growing threat of foodborne bacterial enteropathogens of animal origin. *Clin. Infect. Dis.* **2007**, *45*, 1353–1361. [[CrossRef](#)]
3. Authority, European Food Safety. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2017. *EFSa J.* **2018**, *16*.
4. Adadi, A.; Berrada, M. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160. [[CrossRef](#)]
5. Aruwa, C.E.; Pillay, C.; Nyaga, M.M.; Sabiu, S. Poultry gut health–microbiome functions, environmental impacts, microbiome engineering and advancements in characterization technologies. *J. Anim. Sci. Biotechnol.* **2021**, *12*, 119. [[CrossRef](#)] [[PubMed](#)]
6. Asnicar, F.; Thomas, A.M.; Passerini, A.; Waldron, L.; Segata, N. Machine learning for microbiologists. *Nat. Rev. Microbiol.* **2024**, *22*, 191–205. [[CrossRef](#)]
7. Malakar, S.; Sutaoney, P.; Madhyastha, H.; Shah, K.; Chauhan, N.S.; Banerjee, P. Understanding gut microbiome-based machine learning platforms: A review on therapeutic approaches using deep learning. *Chem. Biol. Drug Des.* **2024**, *103*, e14505. [[CrossRef](#)]
8. Lundberg, S.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**, arXiv:1705.07874v2.
9. Shrikumar, A.; Greenside, P.; Kundaje, A. Learning Important Features Through Propagating Activation. Differences *arXiv* **2019**, arXiv:1704.02685.
10. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.u.; Polosukhin, I. Attention is All you Need. In *Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: San Jose, CA, USA, 2017; Volume 30.
11. Page, L.; Brin, S.; Motwani, R.; Winograd, T. The PageRank Citation Ranking: Bringing Order to the Web. In *Proceedings of the The Web Conference, Toronto, ON, Canada, 11–14 May 1999*.
12. Xu, K.; Ba, J.; Kiros, R.; Cho, K.; Courville, A.; Salakhudinov, R.; Zemel, R.; Bengio, Y. Show, attend and tell: Neural image caption generation with visual attention. In *Proceedings of the International Conference on Machine Learning, PMLR, Lille, France, 6 July–11 July 2015*; pp. 2048–2057.
13. Choi, E.; Bahadori, M.T.; Sun, J.; Kulas, J.; Schuetz, A.; Stewart, W. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. *Adv. Neural Inf. Process. Syst.* **2016**, *29*.
14. Lei, T.; Barzilay, R.; Jaakkola, T. Rationalizing neural predictions. *arXiv* **2016**, arXiv:1606.04155.
15. Jain, S.; Wallace, B.C. Attention is not explanation. *arXiv* **2019**, arXiv:1902.10186.
16. Liu, G.; Zhang, J.; Chan, A.B.; Hsiao, J.H. Human attention-guided explainable artificial intelligence for computer vision models. *Neural Networks* **2024**, *177*, 106392. [[CrossRef](#)] [[PubMed](#)]
17. Kotipalli, B. The Role of Attention Mechanisms in Enhancing Transparency and Interpretability of Neural Network Models in Explainable AI. Master’s Thesis, Harrisburg University of Science and Technology, Harrisburg, PA, USA, 2024.
18. Hwang, D.; Rothrock, M.J.; Pang, H.; Dev Kumar, G.; Mishra, A. Farm management practices that affect the prevalence of Salmonella in pastured poultry farms. *LWT* **2020**, *127*, 109423. [[CrossRef](#)]
19. Rothrock Jr, M.J.; Locatelli, A.; Feye, K.M.; Caudill, A.J.; Guard, J.; Hiatt, K.; Ricke, S.C. A microbiomic analysis of a pasture-raised broiler flock elucidates foodborne pathogen ecology along the farm-to-fork continuum. *Front. Vet. Sci.* **2019**, *6*, 260. [[CrossRef](#)]
20. Xu, X.; Rothrock, M.J.; Mohan, A.; Kumar, G.D.; Mishra, A. Using farm management practices to predict *Campylobacter* prevalence in pastured poultry farms. *Poult. Sci.* **2021**, *100*, 101122. [[CrossRef](#)]
21. Rothrock Jr, M.J.; Hiatt, K.L.; Gamble, J.; Caudill, A.C.; Cicconi-Hogan, K.M.; Caporaso, J.G. A hybrid DNA extraction method for the qualitative and quantitative assessment of bacterial communities from poultry production samples. *J. Vis. Exp. JoVE* **2014**, *94*, e52161.
22. Caporaso, J.G.; Lauber, C.L.; Walters, W.A.; Berg-Lyons, D.; Lozupone, C.A.; Turnbaugh, P.J.; Fierer, N.; Knight, R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4516–4522. [[CrossRef](#)]
23. Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.D.; Costello, E.K.; Fierer, N.; Peña, A.G.; Goodrich, J.K.; Gordon, J.I.; et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **2010**, *7*, 335–336. [[CrossRef](#)]
24. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461. [[CrossRef](#)]
25. DeSantis, T.Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E.L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G.L. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **2006**, *72*, 5069–5072. [[CrossRef](#)]
26. Caporaso, J.G.; Bittinger, K.; Bushman, F.D.; DeSantis, T.Z.; Andersen, G.L.; Knight, R. PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics* **2010**, *26*, 266–267. [[CrossRef](#)] [[PubMed](#)]
27. Huang, X.; Khetan, A.; Cvitkovic, M.; Karnin, Z.S. TabTransformer: Tabular Data Modeling Using Contextual Embeddings. *arXiv* **2020**, arXiv:2012.06678.
28. Von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.* **2007**, *17*, 395–416. [[CrossRef](#)]
29. MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Oakland, CA, USA, 21 June 1967*; Volume 1, pp. 281–297.

30. Jain, A.K.; Murty, M.N.; Flynn, P.J. Data clustering: A review. *ACM Comput. Surv. (CSUR)* **1999**, *31*, 264–323. [[CrossRef](#)]
31. Hoffmann, S.A.; Macculloch, B.; Batz, M. Economic Burden of Major Foodborne Illnesses Acquired in the United States, 2015. Available online: https://www.ers.usda.gov/webdocs/publications/43984/52807_eib140.pdf?v=2344.4 (accessed on 1 June 2024).
32. Vos, P.; Garrity, G.; Jones, D.; Krieg, N.R.; Ludwig, W.; Rainey, F.A.; Schleifer, K.H.; Whitman, W.B. *Bergey's Manual of Systematic Bacteriology: Volume 3: The Firmicutes*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2011; Volume 3.
33. Doyle, M.P.; Diez-Gonzalez, F.; Hill, C. *Food Microbiology: Fundamentals and Frontiers*; John Wiley & Sons: Hoboken, NJ, USA, 2020.
34. Muyyarikkandy, M.S.; Parzygnat, J.; Thakur, S. Uncovering changes in microbiome profiles across commercial and backyard poultry farming systems. *Microbiol. Spectr.* **2023**, *11*, e01682-23. [[CrossRef](#)] [[PubMed](#)]
35. Leser, T.D.; Mølbak, L. Better living through microbial action: The benefits of the mammalian gastrointestinal microbiota on the host. *Environ. Microbiol.* **2009**, *11*, 2194–2206. [[CrossRef](#)] [[PubMed](#)]
36. Liu, L.; Lin, L.; Zheng, L.; Tang, H.; Fan, X.; Xue, N.; Li, M.; Liu, M.; Li, X. Cecal microbiome profile altered by Salmonella enterica, serovar Enteritidis inoculation in chicken. *Gut Pathog.* **2018**, *10*, 34. [[CrossRef](#)]
37. Zhu, J.; Chen, Y.; Imre, K.; Arslan-Acaroz, D.; Istanbulugil, F.R.; Fang, Y.; Ros, G.; Zhu, K.; Acaroz, U. Mechanisms of probiotic Bacillus against enteric bacterial infections. *One Health Adv.* **2023**, *1*, 21. [[CrossRef](#)]
38. Caulier, S.; Nannan, C.; Gillis, A.; Licciardi, F.; Bragard, C.; Mahillon, J. Overview of the antimicrobial compounds produced by members of the Bacillus subtilis group. *Front. Microbiol.* **2019**, *10*, 435128. [[CrossRef](#)]
39. Šimunović, K.; Stefanic, P.; Klančnik, A.; Ereaga, A.; Mandic Mulec, I.; Možina, S.S. Bacillus subtilis PS-216 antagonistic activities against Campylobacter jejuni NCTC 11168 are modulated by temperature, oxygen, and growth medium. *Microorganisms* **2022**, *10*, 289. [[CrossRef](#)] [[PubMed](#)]
40. Luise, D.; Bosi, P.; Raff, L.; Amatucci, L.; Viridis, S.; Trevisi, P. Bacillus spp. probiotic strains as a potential tool for limiting the use of antibiotics, and improving the growth and health of pigs and chickens. *Front. Microbiol.* **2022**, *13*, 801827. [[CrossRef](#)] [[PubMed](#)]
41. Mazanko, M.S.; Popov, I.V.; Prazdnova, E.V.; Refeld, A.G.; Bren, A.B.; Zelenkova, G.A.; Chistyakov, V.A.; Algburi, A.; Weeks, R.M.; Ermakov, A.M.; et al. Beneficial effects of spore-forming Bacillus probiotic bacteria isolated from poultry microbiota on broilers' health, growth performance, and immune system. *Front. Vet. Sci.* **2022**, *9*, 877360. [[CrossRef](#)] [[PubMed](#)]
42. Vivant, A.L.; Garmyn, D.; Piveteau, P. Listeria monocytogenes, a down-to-earth pathogen. *Front. Cell. Infect. Microbiol.* **2013**, *3*, 87. [[CrossRef](#)] [[PubMed](#)]
43. Anjana, A.; Tiwari, S.K. Bacteriocin-producing probiotic lactic acid bacteria in controlling dysbiosis of the gut microbiota. *Front. Cell. Infect. Microbiol.* **2022**, *12*, 851140.
44. Zhang, G.; Raheem, A.; Gao, X.; Zhang, J.; Shi, L.; Wang, M.; Li, M.; Yin, Y.; Li, S.; Cui, X.; et al. Cytoprotective Effects of Lactobacilli on Mouse Epithelial Cells during Salmonella Infection. *Fermentation* **2022**, *8*, 101. [[CrossRef](#)]
45. Webb, L.; Ma, L.; Lu, X. Impact of lactic acid bacteria on the control of Listeria monocytogenes in ready-to-eat foods. *Food Qual. Saf.* **2022**, *6*, fya045. [[CrossRef](#)]
46. Rushdy, A.A.; Goma, E.Z. Antimicrobial compounds produced by probiotic Lactobacillus brevis isolated from dairy products. *Ann. Microbiol.* **2013**, *63*, 81–90. [[CrossRef](#)]
47. Kadam, J.H.; Pawar, R.S.; Din, M.F.M.; Zambare, V. Advances on Probiotics Utilization in Poultry Health and Nutrition. In *Advances in Probiotics for Health and Nutrition*; IntechOpen: London, UK, 2023.
48. Hafez, H.M. Poultry meat and food safety: Pre- and post-harvest approaches to reduce foodborne pathogens. *World's Poult. Sci. J.* **1999**, *55*, 269–280. [[CrossRef](#)]
49. Naureen, Z.; Rehman, N.U.; Hussain, H.; Hussain, J.; Gilani, S.A.; Al Housni, S.K.; Mabood, F.; Khan, A.L.; Farooq, S.; Abbas, G.; et al. Exploring the potentials of Lysinibacillus sphaericus ZA9 for plant growth promotion and biocontrol activities against phytopathogenic fungi. *Front. Microbiol.* **2017**, *8*, 1477. [[CrossRef](#)]
50. Saw, J.H.; Mountain, B.W.; Feng, L.; Omelchenko, M.V.; Hou, S.; Saito, J.A.; Stott, M.B.; Li, D.; Zhao, G.; Wu, J.; et al. Encapsulated in silica: Genome, proteome and physiology of the thermophilic bacterium Anoxybacillus flavithermus WK1. *Genome Biol.* **2008**, *9*, 1–16. [[CrossRef](#)] [[PubMed](#)]
51. Mak, P.H.; Rehman, M.A.; Kiarie, E.G.; Topp, E.; Diarra, M.S. Production systems and important antimicrobial resistant-pathogenic bacteria in poultry: A review. *J. Anim. Sci. Biotechnol.* **2022**, *13*, 148. [[CrossRef](#)] [[PubMed](#)]
52. Sanjorjo, R.A.; Tseten, T.; Kang, M.K.; Kwon, M.; Kim, S.W. In Pursuit of Understanding the Rumen Microbiome. *Fermentation* **2023**, *9*, 114. [[CrossRef](#)]
53. Fusco, W.; Lorenzo, M.B.; Cintoni, M.; Porcari, S.; Rinninella, E.; Kaitsas, F.; Lener, E.; Mele, M.C.; Gasbarrini, A.; Collado, M.C.; et al. Short-Chain Fatty-Acid-Producing Bacteria: Key Components of the Human Gut Microbiota. *Nutrients* **2023**, *15*, 2211. [[CrossRef](#)] [[PubMed](#)]
54. Zhao, Y.; Wei, H.M.; Yuan, J.L.; Xu, L.; Sun, J.Q. A comprehensive genomic analysis provides insights on the high environmental adaptability of Acinetobacter strains. *Front. Microbiol.* **2023**, *14*, 1177951. [[CrossRef](#)] [[PubMed](#)]
55. Davies, A.R.; Board, R.; Board, R. *Microbiology of Meat and Poultry*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1998.
56. Joshi, S.G.; Litake, G.M. Acinetobacter baumannii: An emerging pathogenic threat to public health. *World J. Clin. Infect. Dis.* **2013**, *3*, 25–36. [[CrossRef](#)]

57. Samtiya, M.; Matthews, K.R.; Dhewa, T.; Puniya, A.K. Antimicrobial resistance in the food chain: Trends, mechanisms, pathways, and possible regulation strategies. *Foods* **2022**, *11*, 2966. [[CrossRef](#)] [[PubMed](#)]
58. Calderon, R.B.; Jeong, C.; Ku, H.H.; Coghill, L.M.; Ju, Y.J.; Kim, N.; Ham, J.H. Changes in the microbial community in soybean plots treated with biochar and poultry litter. *Agronomy* **2021**, *11*, 1428. [[CrossRef](#)]
59. Shah, D.H.; Board, M.M.; Crespo, R.; Guard, J.; Paul, N.C.; Faux, C. The occurrence of Salmonella, extended-spectrum β -lactamase producing Escherichia coli and carbapenem resistant non-fermenting Gram-negative bacteria in a backyard poultry flock environment. *Zoonoses Public Health* **2020**, *67*, 742–753. [[CrossRef](#)]
60. Herrera, P.; Kwon, Y.M.; Ricke, S.C. Ecology and pathogenicity of gastrointestinal Streptococcus bovis. *Anaerobe* **2009**, *15*, 44–54. [[CrossRef](#)]
61. Hyun, D.W.; Whon, T.W.; Cho, Y.J.; Chun, J.; Kim, M.S.; Jung, M.J.; Shin, N.R.; Kim, J.Y.; Kim, P.S.; Yun, J.H.; et al. Genome sequence of the moderately halophilic bacterium Salinicoccus carniancri type strain Crm T (=DSM 23852 T). *Stand. Genom. Sci.* **2013**, *8*, 255–263. [[CrossRef](#)] [[PubMed](#)]
62. Cycil, L.M.; DasSarma, S.; Pecher, W.; McDonald, R.; AbdulSalam, M.; Hasan, F. Metagenomic insights into the diversity of halophilic microorganisms indigenous to the Karak Salt Mine, Pakistan. *Front. Microbiol.* **2020**, *11*, 1567. [[CrossRef](#)] [[PubMed](#)]
63. Klein, G. Taxonomy, ecology and antibiotic resistance of enterococci from food and the gastro-intestinal tract. *Int. J. Food Microbiol.* **2003**, *88*, 123–131. [[CrossRef](#)] [[PubMed](#)]
64. Micallef, S.A.; Goldstein, R.E.R.; George, A.; Ewing, L.; Tall, B.D.; Boyer, M.S.; Joseph, S.W.; Sapkota, A.R. Diversity, distribution and antibiotic resistance of *Enterococcus* spp. recovered from tomatoes, leaves, water and soil on US Mid-Atlantic farms. *Food Microbiol.* **2013**, *36*, 465–474. [[CrossRef](#)]
65. Bao, Y.; Dolfing, J.; Guo, Z.; Chen, R.; Wu, M.; Li, Z.; Lin, X.; Feng, Y. Important ecophysiological roles of non-dominant Actinobacteria in plant residue decomposition, especially in less fertile soils. *Microbiome* **2021**, *9*, 84. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.