Physical Education. Health and Social Sciences
https://journal-of-social-education.org
E-ISSN: 2958-5996
P-ISSN: 2958-5988

Early Detection and Prediction of Zoonotic Virus Integrated with Machine Learning Techniques

Asma Kiran Pervez[1*], Muhammad Adeel Anjum[2], Muhammad Waqas Riaz[3], Muhammad Aftab[4], Muhammad Yousuf[5], Muhammad Yousif[6]

[1] National University of Modern Languages, Lahore Campus & Ikram ul Haq, Institute of Industrial Biotechnology, Govt College University Lahore
[2,6] Department of Computer Science, National University of Modern Languages, Lahore campus, Lahore, Pakistan
[3,4] Department of Artificial Intelligence, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan
[5] Faculty of Computer Science, Minhaj University, Lahore, Pakistan
Email: [1]akpervaiz@numl.edu.pk, [2]adeel.anjum701@gmail.com, [3]mwaqaskp@gmail.com, [4] Chaudharymaftab@gmail.com, [5] 2024f-mulms-ds-009@mul.edu.pk, [6] myousif.cs@gmail.com
Corresponding Author Email: asmakiranpervez@gmail.com

DOI:* *https://doi.org/10.63163/jpehss.v4i1.1304

Abstract:
Zoonotic virus, which animals transmit to people, remain an increasing threat to human health in the world due to their capacity to propagate rapidly and evolve over time. This paper examines how machine learning can be applied to aid early diagnoses and anticipation of these diseases. It talks about the various forms of infectious entities such as viruses, bacteria, fungi and parasites and the use of animals as carriers. The recent outbreaks have been also reviewed in the paper and the role of learning about the spread of these diseases and the factors that increase the risks have been highlighted. The prediction framework is provided which utilizes the real-life data obtained both in medical and environmental sphere. Data cleaning and preparation are done to be divided into training and testing data. A number of machine learning models, including Decision Tree, KNN, SVM, Logistic Regression, Random Forest, and XGBoost are used and compared. Their output is measured in standard measures in order to estimate which model is the best one in predicting potential infections. The results indicate that the combination of models and the more sophisticated techniques may contribute to the high quality of the predictions and the possibility to take the timely preventive measures. Overall, this study supports the use of data-driven methods to strengthen disease monitoring systems and reduce the impact of zoonotic outbreaks.

Keywords: Zoonotic Virus, Animal to human, ML Techniques, Machine Learning

Introduction:

Zoonotic virus which spreads from animal to human so vertebrates are animals which have backbone, like cow, sheep, rats, dogs, cats, bats and birds. The living style of human and animals have both influence and germs spread in the everywhere air, soil and water. Germs are part of life some germs are harmful other are good for health. Ebola virus feels fever, sore throat, bleeding and organ failure. Numerous infections only affect one particular kind of organism, such as people, certain animals, plants, or even other microorganisms [1]. However, humans and other vertebrates can contract zoonotic diseases. Alternatively, they used to solely infect particular animals, but mutations made it possible for them to "jump" to people and spread the illness. Germs have 4 types one of them are bacteria, viruses, fungi and protozoa [2]. Bacteria are single-celled, microscopic living things. "Bacteria" is the word for just one. The planet is home to millions, if not billions, of different kinds of bacteria, including those in your body. They are in your mouth, airways, and on your skin. Additionally, they are found in your urinary tract, reproductive system, and digestive system. According to scientists, your body has ten times as many bacterial cells as human cells [3]. Microscopic creatures known as viruses have the ability to infect humans, plants, and animals. They are a tiny fragment of genetic material (DNA or RNA) enclosed in a capsid for protection. Additionally, some viruses have an envelope. Viruses require a host in order to reproduce. The flu, the common cold, and COVID-19 are a few prevalent viral infections [4]. Any illness or ailment contracted from a **fungus** is referred to as a fungal infection. Although they can infect your lungs or other regions of your body, they typically affect your skin, hair, nails, or mucous membranes. You are more susceptible to fungal infections if your immune system is compromised. Fungal infections are typically treated with antifungal drugs [5].

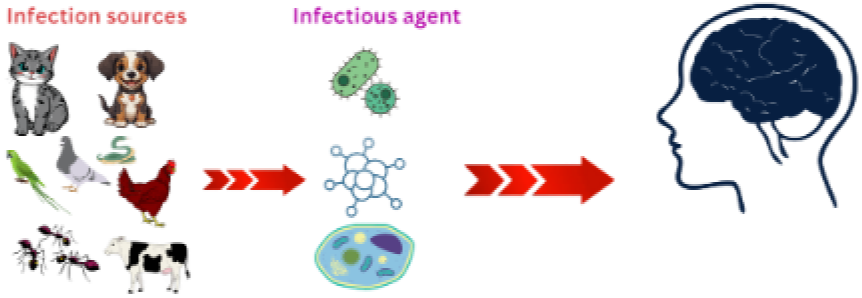

Figure 1. Illustration of the concept of infectious agent transmission mechanisms to Human.

A class of uncommon, fatal neurodegenerative illnesses is known as prism disease. They occur when proteins in your brain change into prions, which are aberrant proteins. Dementia results from brain damage brought on by prism illness. Medical professionals concentrate on symptom management strategies, such as medication [6]. Influenza A kind of virus A subtype of the influenza A virus is known as **H1N2**. Nowadays, it is endemic in pig populations, while humans might also get it sometimes. No notable increases in influenza activity have been linked to the virus, and it does not produce more severe illness than other influenza viruses [7]. Influenza A kind of virus the influenza A virus has a subtype known as **H2N2**[8]. The "Asian flu" strain, **H3N2**[9], and other strains identified in birds are among the strains of H2N2 that have undergone mutations.

Table 1. Influenza A subtypes

Subtype	**Primary Hosts**	**Human Infection**	**Notes / Impact**
H1N1	Humans, pigs, birds	Yes	Caused 1918 Spanish Flu (50M deaths), 2009 Swine Flu pandemic. Still circulating seasonally.
H2N2	Birds, humans (historic)	Yes (past)	Caused 1957 Asian Flu pandemic. No longer circulating in humans.
H3N2	Humans, birds, pigs, dogs	Yes	Caused 1968 Hong Kong Flu pandemic; still a seasonal flu strain.
H5N1	Birds (especially poultry), cats, humans (rare)	Yes (sporadic, high mortality)	Highly pathogenic avian influenza (HPAI). Human cases since 1997, mortality >50%.

H7N9	Birds (esp. poultry)	Yes (sporadic, severe)	Emerged in China (2013). High mortality, pandemic potential.
H9N2	Birds (poultry), pigs	Yes (mild cases)	Common in poultry, occasional mild human infections. Acts as a donor of internal genes to other subtypes.
H10N8	Birds	Yes (rare)	Few human cases in China, 2013–2014, some fatal.
H1N2	Humans, pigs	Yes (sporadic)	Circulates in swine and occasionally infects humans.
H3N8	Horses, dogs, birds	Rare	Recently (2022) reported first human infection in China. Potential zoonotic concern.
H7N7	Birds, horses	Yes (rare)	One human fatality reported in the Netherlands (2003 poultry outbreak).

Outbreaks of Zoonotic Virus

Table 2.Timeline of Recent Zoonotic Outbreaks

Year	Disease / Outbreak	Notes (Location & Significance)	Main Transmission Route
2020	COVID-19 (SARS-CoV-2)	Global pandemic; likely zoonotic origin with sustained human-to-human spread	Initially animal-to-human (suspected), now mainly respiratory droplets/aerosols between people
2021	Foodborne zoonoses (e.g., Salmonella, Campylobacter) and COVID-19 variants (Delta)	Common reported zoonotic diseases; COVID-19 Delta wave dominated	Foodborne (contaminated meat/eggs) for Salmonella/Campylobacter; respiratory person-to-person for COVID-19
2022	Mpox (Monkeypox virus)	Global spread beyond endemic African countries	Close contact with lesions/body fluids, respiratory droplets, or contaminated materials; also animal-to-human from infected animals
2023	Avian Influenza (H5N1)	Human cases (e.g., Cambodia); continued bird outbreaks worldwide	Direct contact with infected birds or contaminated environments; rare human-to-human
2024	Avian Influenza (H5N1) and Mpox (Clade I)	Human H5N1 case in USA linked to animal exposure; Mpox resurgence in Africa	H5N1 contact with infected birds/mammals Mpox close contact with infected persons/animals
2025	Marburg virus (suspected clusters) and ongoing H5N1 & Mpox monitoring	Suspected Marburg clusters (e.g., Tanzania); continued surveillance of other zoonoses	Marburg spillover from bats, then direct human-to-human via fluids/contact; H5N1 & Mpox routes as above

Related Work

The transmission dynamics of rabies involve the spread of the virus within populations of animals and in some cases from animals to humans. Severity of disease is dependent on different aspects like species involved, geographical locality, and control measures in place [10]. Certain animal species, known as reservoir hosts are the main reason to affect the spread and transmission of the virus in the certain geographic region. Domestic dogs, for example, are the principal reservoir host for rabies in many regions of the world and contribute to the transmission cycle [11]. Wildlife such as bats, raccoons, foxes and skunks act as reservoir hosts for rabies in many areas. Transmission among wildlife populations can happen through bites but it can also happen through other means, such as contact with infected surfaces or inhalation of aerosolized virus in bat roosts [12]. Rabies is typically transmitted to animals through the bite of an infected animal. The virus can be transferred if a rabies infested animal bites another vulnerable animal. This transmission can occur within a species or across species [13]. Although it is uncommon but human-to-human rabies transmission can occur by organ donation from infected donors or through extremely intimate contact, such as bites or exposure to contaminated saliva [12]. Following transmission, the virus replicates within the host's body and transfers to the nervous system from the peripheral parts. The time frame and the duration of the incubation period can vary from several weeks to months [13]. Severe fever with thrombocytopenia syndrome (SFTS) is a tick-borne viral disease first reported in China in 2011 and later in Japan in 2013. It causes fever, fatigue, low white blood cells, low platelets, and sometimes severe complications, with a fatality rate of around 16% in Japan. While ticks are the main source of transmission, cases of human-to-human and animal-to-human spread have also been reported. The article describes the first confirmed cat-to-human transmission of SFTSV in Japan. A veterinarian became infected after caring for and performing autopsies on sick cats, despite using partial protective equipment. Genetic testing confirmed the virus in both the patient and the cats was identical. This case shows that cats can be a direct source of infection and highlights the importance of strict protective measures, especially for veterinarians and others working with sick animals [14]. Sporotrichosis is a fungal infection mainly caused by Spoorthi Brasiliense's. Once linked to soil and plants, it is now increasingly transmitted from cats to humans through scratches, bites, or lesions. Brazil, especially Rio de Janeiro, has seen a large epidemic, while smaller outbreaks appear in other countries. Treatment is long and difficult due to drug resistance, and new therapies and vaccines are under study. Overall, cat-transmitted sporotrichosis is an emerging zoonotic threat needing strong public health and veterinary action [15]. Severe fever with thrombocytopenia syndrome (SFTS) is an emerging tick-borne viral disease first identified in China in 2011 and later reported in Japan and Korea, with fatality rates of 6–30%. It causes fever, leukopenia, thrombocytopenia, and multi-organ dysfunction. While ticks are the main vectors, human-to-human and animal-to-human transmissions have been documented, with cats posing a particularly high zoonotic risk due to severe illness and high viral loads. No specific treatment or vaccine exists, though antivirals and vaccines are under study. SFTS is therefore a growing public health and occupational threat, demanding coordinated veterinary and medical strategies for prevention and control [16].

Proposed Methodology

We discuss some machine learning research questions that may be important to zoonotic emerging disease detection. When appropriate, we suggest potential ap preaches to address each question. Our discussion is aimed at improving awareness in both the machine learning [17] and wildlife monitoring communities, rather than offering definite solutions.

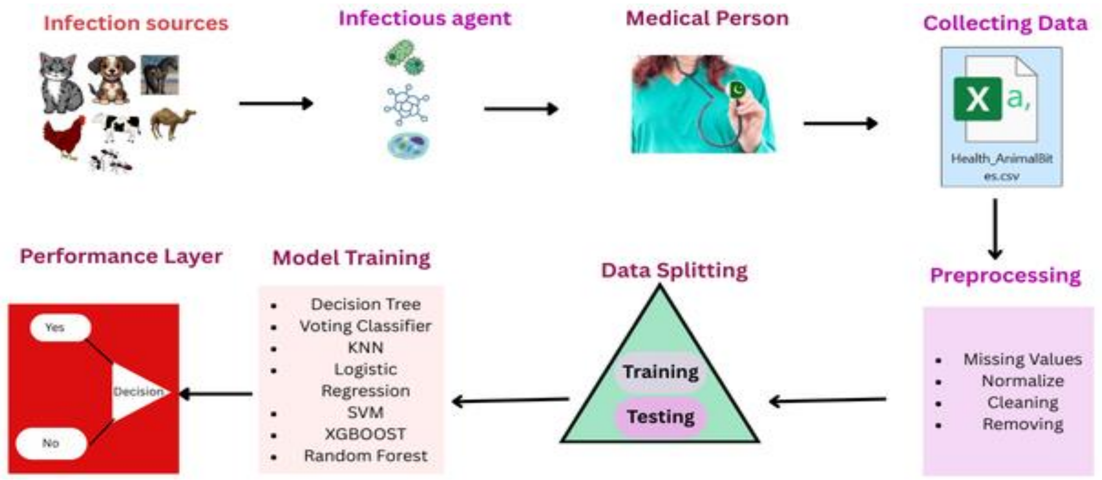

Figure 2. Workflow of Proposed Research

Infection Sources:

Figure 1 shows a complete workflow for predicting zoonotic diseases using machine learning. It begins with infection sources (animals) that carry infectious agents, which are detected and recorded by medical professionals. The collected health data is stored in files and then preprocessed to handle missing values, clean and normalize the data. After preprocessing, the data is split into training and testing sets, and various machine learning models such as Decision Tree, KNN, SVM, Logistic Regression, XGBoost, and Random Forest are trained on the training data. Finally, the trained models are evaluated in the performance layer to make decisions (yes/no) about the presence or risk of infection, supporting early detection and disease prediction.

Infection Sources:

Infection Sources are the original carriers or reservoirs of infectious agents (such as viruses, bacteria, or parasites) that can spread diseases to humans. In the diagram, animals like dogs, cats, cows, chickens, and horses are shown as these sources. They may carry zoonotic pathogens either without showing symptoms (acting as reservoirs) or while being sick themselves. When humans come into contact with these animals through bites, scratches, handling, or consuming their products these infectious agents can be transmitted and cause disease.

Infection Agent:

These are the **microorganisms that actually cause diseases** when they enter a host's body. In the diagram, they are shown as viruses, bacteria, and other microscopic particles. These agents originate from infection sources (like animals) and can spread to humans, leading to zoonotic diseases. Common types include:

- **Viruses** (e.g., rabies virus, influenza virus)
- **Bacteria** (e.g., Salmonella, Brucella)
- **Parasites** (e.g., Toxoplasma)
- **Fungi** (e.g., ringworm-causing fungi)

They are the key link between animal sources and human infections without the infectious agent, disease transmission cannot occur.

Medical Person

A medical professional (doctor, veterinarian, or health worker) detects and records the occurrence of infections in humans and animals. They serve as the link between the infected cases and the data collection process.

Collecting Data:

Collecting data for zoonotic viruses involves gathering health-related information from both animals and humans to track and analyze how these viruses spread. This data is usually collected by medical or veterinary professionals from various sources [18], such as:

- **Animal records:** information on sick or dead animals, animal bite incidents, vaccination history, and species involved.
- **Human health records:** patient symptoms, lab test results, travel or exposure history, and confirmed infection cases.
- **Laboratory data:** test results confirming the presence of a specific virus in samples from animals or humans.
- **Environmental and location data:** where and when the cases occurred, climate conditions, and animal movement patterns.

All this information is compiled into datasets (often in formats like CSV or Excel files) and later used for analysis, prediction, and early detection of zoonotic virus outbreaks using machine learning models.

Preprocessing:

It is the step of preparing raw collected data for analysis by cleaning and organizing it so machine learning models can use it effectively. This is used in the zoonotic virus prediction to eliminate duplicate or irrelevant entries, process missing values, transform text or categorical data (such as animal type or location) to numeric data, normalize values to a comparable scale, and balance between the number of infected and non-infected cases. This will remove duplications, errors, and inaccurate data in the data, thereby enhancing accuracy and reliability of the prediction models.

Data Splitting:

It consists of splitting the ready-made set into two groups: a training set and a testing set. The machine learning model is trained on the training set to learn to identify patterns, segmentation[19] and relationships in the data, and tested on the testing set to determine the effectiveness of the model on unseen data. This assists in making sure that the model will learn and not simply memorize the data (overfitting) and be able to predict the results of new cases correctly, including detecting possible zoonotic virus infections.

Model Training:

Decision Tree: Constructs a tree of rules by dividing data by feature values to make decisions (easy to understand).

Voting Classifier: The voting Classifier takes the results of several different models and majority of the models is used to make the final decision, which has better accuracy.

k-Nearest Neighbors (KNN): Classifies a case by using the labels of the nearest data points (neighbors) in the data set.

Logistic Regression: A statistical model which estimates the likelihood of an outcome (infected or not) given input features.

Support Vector Machine (SVM): Identifies the optimal boundary (hyperplane) between the various classes (infected vs non-infected).

XGBoost: The XGBoost is a powerful boosting algorithm that creates many small decision trees sequentially to enhance accuracy and minimise errors.

Random Forest: Establishes several decision trees on random data sets and pools the outcomes to provide more consistent and precise forecasts.

Performance Layer:

The last step is the performance layer, in which trained machine learning models are tested on unknown data to determine its capacity to predict the results accurately. It evaluates the model on metrics such as accuracy, precision, recall, F1-score, and confusion matrix to assess the model in identifying infected and non-infected cases. On the basis of these results, the most precise one is chosen and in case its performance is credible, it may be applied to make real-time yes/no decisions regarding the presence or threat of zoonotic infections.

Simulation and Results:

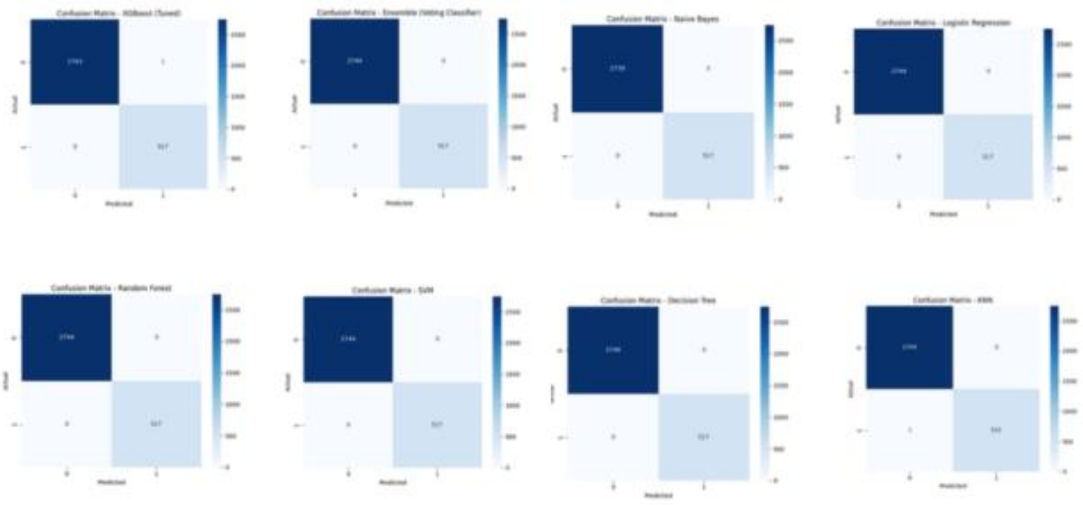

Figure 3. Confusion Matrix for Machine learning Techniques

Figure 3 shows the confusion matrices of eight machine learning models used to do classification: Gradient Boosting (Tuned), Ensemble (Using Classifier), Naive Bayes, Logistic Regression, Random Forest, SVM, Decision Tree, and KNN. The number of true positives, false positives, false negatives, and true negatives are demonstrated in each of the matrices, and they assist in evaluating the models in terms of their ability to differentiate true and predicted classes. The more the values along the diagonal (true positives and true negatives), the more accurate model predictions are made. Gradient Boosting and Ensemble are more likely to perform better with few misclassifications whereas KNN has more false positives and false negatives meaning more misclassifications and less accurate.

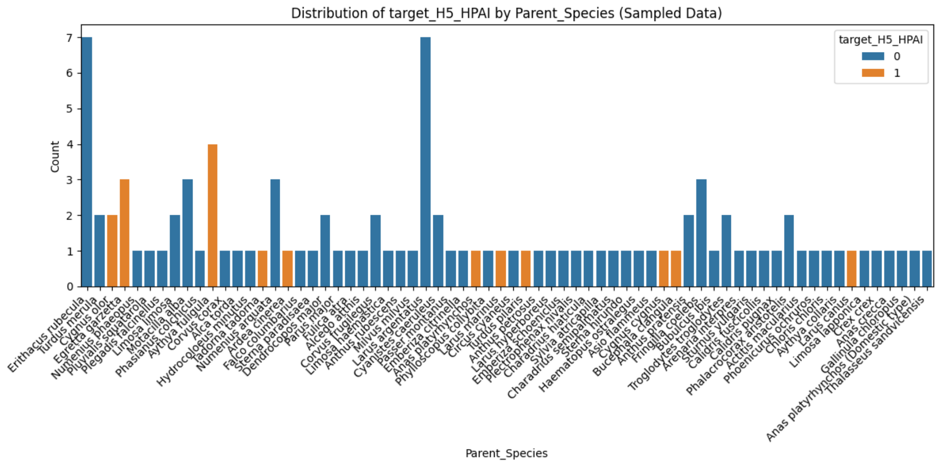

Figure 4. Distribution of all dataset of Type virus

Figure 4 depicts the distribution of the variable target H5 HPAI by Parent Species. The plot is used to show the number of each species, targeting H5HPAI = 0 (blue) and targeting H5HPAI = 1 (orange). Parent_Species categories are represented on the x-axis, and the number of occurrences of each species on the y-axis. Based on the plot, it is clear that some species, such as Erythrosuchus, possess a significantly greater number of target_H5_HPAI = 0, which means that there are many non-H5 HPAI cases. Conversely, such species as Anas platyrhynchos and Phalacrocorax carbo exhibit the balance between target_H5_HPAI = 0 and target_H5_HPAI = 1, suggesting that there are infected and non-infected individuals. The visualization can help to comprehend the distribution of H5 HPAI infection among various species within the dataset.

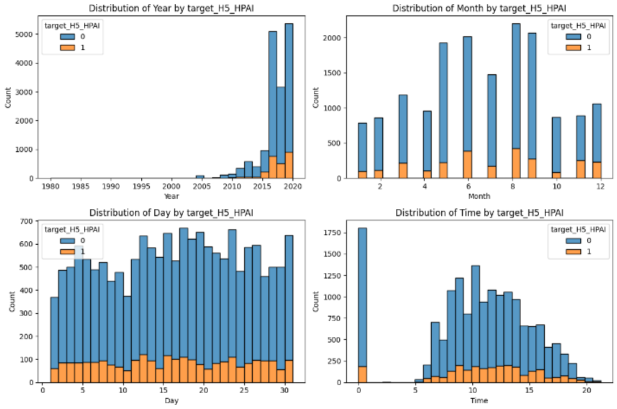

Figure 5. Distribution by variables

Figure 5 shows the distribution of target_H5_HPAI by different time based classes: Year, Month, Day, and Time. The Year plot shows a sharp rise in target_H5_HPAI = 1 (orange) since 2015, with a surge in cases occurring around 2020, implying that there has been a strong increase in the cases of H5 HPAI over the years. Month plot This indicates a fairly balanced distribution of cases across the year with particular months experiencing higher prevalence of target_H5_HPAI = 1, especially in the month of December. The Day plot shows a more stable distribution where a small rise in the cases of target H5 HPAI = 0 (blue) in the middle of the month, but target H5 HPAI = 1 is well spread. Finally, the Time plot depicts that H5 HPAI cases are clustered within certain hours of the day and especially early morning and late evening. These plots combined indicate that there is an increasing trend of H5 HPAI over time with some seasonal and daily patterns that may be significant to the spread and risk of the disease.

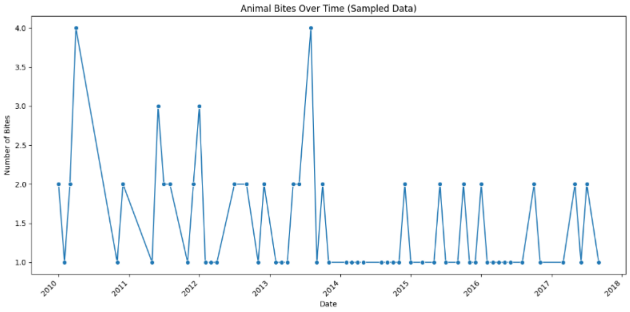

Figure 6. Animal Bites Over Time

The figure 6 shows the number of animal bites per year between 2010 and 2018. The line chart indicates that there have been large variations in the number of bites, with some spikes, as they occurred in 2014, when the number of bites grew to almost 4. The range of the data has the frequent years with lower numbers implying the relative stability or minimum incidents, whereas particular years, such as 2013 and 2014, demonstrate significant peaks. These spikes can be in the form of certain events or times of activities or activities that involved animals. Altogether, the data show a tendency of irregular cases without a regular tendency, but it shows individual years with the highest cases of bites, which might be explained by some other factors, such as weather or environmental conditions.

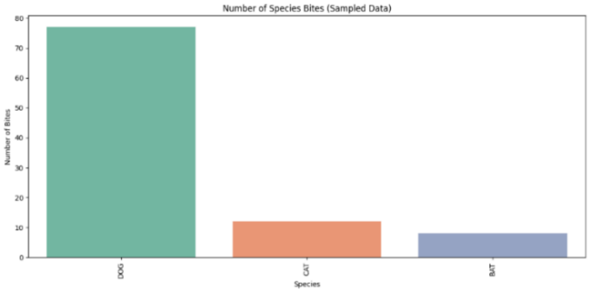

Figure 7. Number of Species Bites

Figure 7 shows the species bites of three categories, Dog, Cat and Bat. The bar chart evidently demonstrates that dogs are the most common source of bites with more than 70 dogs bites and a very low number of cat and bat bites (approximately 20 and 10, respectively). This is an indication that dogs are the leading species causing animal bites in the data sampled with cats and bats having a relatively less number of cases. The chart gives a clear graphical comparison of the incidents of biting by the species, and it is evident that in this case, dogs dominate.

Conclusion

The aim of this literature survey is to synthesize and analyze machine learning and deep learning approaches applied to study zoonotic diseases. Our review findings will enable researchers to understand predictive models to identify the risk factors for transmission to develop mitigation strategies. The survey revealed that traditional machine learning models are widely used in this field. According to our findings, support vector machines are flexible enough to learn population growth dynamics and predict the occurrence of diseases. With noisy, complex, and varied data, ensemble approaches such as random forest and xgboost have demonstrated excellent performance. However, deep learning methods have tremendous potential for identifying appropriate protective models. Application of deep learning techniques, such as segmentation and classification of images, can enhance research into diagnosing irregularities caused by infections. While the resources in this field are limited, transfer learning, where we reuse a previously trained model as the basis for training a new model, or zero-shot-based learning that classifies data based on very few or even no labeled examples, have the potential to make learning more efficient and contribute to the development of diagnostic and preventive strategies to limit the spread of zoonotic diseases.

References:

[1] Yousaf M et al., 2012. Rabies molecular virology, diagnosis, prevention and treatment. Virology Journal 9: 1-5

[2] Coetzer A et al., 2019. Epidemiological aspects of the persistent transmission of rabies during an outbreak (2010 2017) in Harare, Zimbabwe. PLoS One 14: 0210018.

[3] Lembo T et al., 2008. Exploring reservoir dynamics: a case study of rabies in the Serengeti ecosystem. Journal of Applied Ecology 45: 1246-1257.

[4] Lushasi K et al., 2021. Reservoir dynamics of rabies in south-east Tanzania and the roles of cross-species transmission and domestic dog vaccination. Journal of Applied Ecology 58: 2673-2685.

[5] Rupprecht CE et al., 2017. Lyssaviruses and rabies: current conundrums, concerns, contradictions and controversies. F1000Research 6: 28299201.

[6] Kida, K., Matsuoka, Y., Shimoda, T., Matsuoka, H., Yamada, H., Saito, T., ... & Kishimoto, T. (2019). A case of cat-to-human transmission of severe fever with thrombocytopenia syndrome virus. *Japanese Journal of Infectious Diseases*, *72*(5), 356-358.

[7] Frymus, T., Belák, S., Egberink, H., Hofmann-Lehmann, R., Marsilio, F., Addie, D. D., ... & Hosie, M. J. (2021). Influenza virus infections in cats. Viruses, 13(8), 1435.

[8] Li, R., Han, Q., Li, X., Liu, X., & Jiao, W. (2024). Natural product-derived phytochemicals for influenza A virus (H1N1) prevention and treatment. *Molecules*, *29*(10), 2371.

[9] Zambon, M., & Hayden, F. G. (2026). Influenza A (H3N2) subclade K virus: threat and response. *JAMA*, *335*(4), 307-310.

[10] Fatima, M., Iqbal, T., Shaheen, L., Salma, U., Siddique, R., Ali, R., ... & Usman, S. (2023). Transmission dynamics of rabies virus. *Zoonosis, Unique Scientific Publishers, Faisalabad, Pakistan*, *3*, 386-397.

[11] Escobar, L. E., Velasco-Villa, A., Satheshkumar, P. S., Nakazawa, Y., & Van de Vuurst, P. (2023). Revealing the complexity of vampire bat rabies" spillover transmission". *Infectious diseases of poverty*, *12*(01), 102-110.

[12] Pittman, T. W., Decsi, D. B., Punyadeera, C., & Henry, C. S. (2023). Saliva-based microfluidic point-of-care diagnostic. *Theranostics*, *13*(3), 1091.

[13] Lushasi, K., Brunker, K., Rajeev, M., Ferguson, E. A., Jaswant, G., Baker, L., ... & Hampson, K. (2022). Integrating contact tracing and whole-genome sequencing to track the elimination of dog-mediated rabies. *medRxiv*, 2022-11.

[14] Pasfield, K., Gottlieb, T., Tartari, E., Ward, M. P., & Quain, A. (2022). Sickness presenteeism associated with influenza-like illness in veterinarians working in New South Wales: Results of a state-wide survey. *Australian veterinary journal*, *100*(6), 243-253.

[15] de Andrade Galliano Daros Bastos, F., de Farias, M. R., Gremião, I. D. F., Bridi Cavassin, F., dos Santos Monti, F., Vilas Boas, R. R., ... & Queiroz-Telles, F. (2025). Cat-transmitted sporotrichosis by Sporothrix brasiliensis: Focus on its potential transmission routes and epidemiological profile. *Medical Mycology*, *63*(6), myaf051.

[16] Ebadi, A. G., & Selamoglu, Z. (2025). Advances in Veterinary Medicine: Sustainable Approaches to Animal Health Man-agement and Disease Prevention. *J Anim Biol Vet*, *4*, 1-13.

[17] Yousif, M., Abbas, A., Hasan, Z., Ali, D., & Sarfraz, M. (2021). Smart Village Health System IoT to Envisage Chronical Disease Using Artificial Neural Network. *Internet Things Cloud Comput.*, *9*(4), 27.

[18] https://www.kaggle.com/datasets/shijo96john/animal-disease-prediction

[19] Yousif, M., Rehman, F., Nagra, A. A., Saleem, S., Abdulwahid, A., & Sarfraz, M. (2025). Color Image Segmentation Optimization: Threshold Edge Detection with Harmonic and Wiener Filter Enhancements. *Journal of Computing & Biomedical Informatics*, *9*(01).