Advances in Machine Learning for Predicting and Detecting Influenza Outbreaks: A Review

Ehsan khodadadi[1,*]

[1]Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR 72701, USA

Emails: Ehsank@uark.edu

Abstract

Influenza is often associated with millions of cases and hundreds of thousands of deaths each year, thus constituting a serious threat to public health. Traditional surveillance techniques employed in epidemiology are limited in forecasting impending outbreaks as caused by delays in receiving the relevant information and the dynamic nature of political environments. This review focuses on the available literature on the use of machine learning (ML) techniques in understanding and controlling influenza with an accent on all the sources of information available, including clinical papers, social networking sites and others. Applicable practices in classifying predictive modeling techniques, including deep learning and others, ensemble techniques, time series analysis, etc., have increased the speed and precision of the earlier results. Even so, the achievements made so far have not come on a silver platter as there are challenges, but not limited to data issues, model explain ability and strict validation processes. Some research areas are enhancing the present models to accommodate diverse virulent strains of the viruses and advancing extensive data analysis methods. It is noted in this review that machine learning strategies are essential in combating health issues and, thus, why such technologies can be deployed within a concise duration in the context of influenza epidemics for effective forecasting and resource management to salvage lives.

Keywords: Influenza prediction; Machine learning; Outbreak detection; public health surveillance; Deep Learning

1. Introduction

This paper segment evaluates resources on different components and modeling processes employed in predicting influenza. Influenza - an infectious illness of the respiratory system - has claimed and continues to claim a considerable number of victims around the world [1]. For this reason, studying incidence patterns and predicting the future threats of influenza infection becomes essential for developing, planning, and implementing effective preventative measures such as outbreak control, optimal resource allocation and timely vaccine provision [2]. The scope of this review encompasses several research works that have looked at different techniques for influenza prediction, such as statistics, machine learning, data mining, and analysis of social networks. Given these concerns, we devote ourselves to advancing and researching the predictions of influenzas [3].

Classic approaches to anticipate flu outbreaks most often include existing statistical evaluations, which also consist of epidemiological and clinical data. Although efficient, these techniques are prone to challenges such as the time it takes to report relevant information and the inability to cope with dynamic shifts in behavior regarding viruses spreading. Unlike Twitter messaging, which is often instantaneously available following a user's post, traditional surveillance systems suffer from a delay between 1-2 weeks from when outbreaks begin to when data are available [4].

In recent years, there has been increasing attention to using machine learning (ML) methods to improve influenza prediction models. Because ML approaches span various algorithms, from supervised learning to deep learning [5], they can process big data and sift even the intricate relationships that statistical techniques might overlook. They also enjoy being trained on different datasets, such as patient information, activity on social media, and weather patterns, among others, for better and faster predictions [6].

DOI: https://doi.org/10.54216/JAIM.080204
Received: March 28, 2024 Revised: June 05, 2024 Accepted: November 08 2024

Machine learning has the advantage of simultaneously working with large volumes of data from several sources when predicting influenza epidemics. For instance, data available on social networks, web searches, and news can be synthesized to analyze people's activities and engagement, which can be precursors to even the earliest stages of an epidemic. Such data sources with ML techniques have been proven to increase foresight in disease outbreak prediction [7].

Pneumonia forecasting has been attempted through various machine learning methods, including convolutional neural networks (CNNs), support vector machines (SVMs) and their composites. Every method has its advantages, such as how CNNs can utilize images for pattern research or how ensemble models are suitable for working with mixtures of data. These techniques provide considerably better results than classical models in estimating when and where the spread of an influenza infection will occur [8].

Influenza research has also benefited from deep learning, a subfield of machine learning. To illustrate, deep convolutional neural networks (CNN) have been able to predict the antigenic properties of different strains of influenza, which aids in vaccine development. These models have been proven to be very precise, so it can be argued that they are valuable in predicting which strains will be most widespread in the coming seasons, as their information will thus be helpful in vaccine development [9].

Besides their predictive power, machine learning models are scalable and flexible, so they can be used to track influenza disease patterns worldwide. Advanced models analyze spatiotemporal data to project the dynamics of influenza in different regions, factoring in climate, population density, and people's movement. For example, the combination of machine learning (ML) techniques and geographic information systems (GIS) has helped map the ecological ranges of different influenza virus variants and their transmission dynamics [10].

However, several barriers that must be addressed hinder the effective use of machine learning models for predicting influenza. These include the data quality, the models' explain ability, and the necessity for extreme cross-validation. Moreover, the complexity is compounded by the demand for accurate and current information from diverse health facilities, networks, and health systems-oriented databases within a short time frame [11].

Contemporary research will focus on implementing more advanced data analytical techniques, making machine learning algorithms interpretable, and including newly available information to improve forecasting accuracy. For instance, it promotes the potential for pandemic readiness and active monitoring mechanisms using graph-based approaches sensitive to pattern changes over time. Advancements in these technologies may change public health officials' approach to managing Ebola Virus Disease outbreaks [12], [13].

This paper analyzes advanced machine-learning methods' comparative ease and effectiveness for forecasting and detecting influenza outbreaks. We intend to present a balanced picture of the current state of research on history-based approaches by analyzing its recognized effects, challenges and prospects. The importance of understanding the advantages and limitations of these methods becomes evident when it is seen in the context of the goal to improve early-warning systems and when, in turn, this goal is seen as part of public health policy aimed at controlling epidemics of influenza.

2. Literature Review

This section of the paper reviews the existing literature on the various factors and modeling approaches of influenza prediction. Influenza, a contagious disease of the respiratory system, has been a bane to many people worldwide. Therefore, understanding the incidence patterns and future risks of influenza infection becomes imperative in implementing timely and effective public health strategies, namely controlling the outbreaks, efficiently managing resources and optimizing vaccine development. The scope of this review encompasses several research works that have looked at different techniques for influenza prediction, such as statistics, machine learning, data mining, and analysis of social networks. Considering these issues, we study developments in these directions and summarize advances in research on influenza predictions.

According to the paper described above [14], evaluating rupture potential in patients with abdominal aortic aneurysms (AAA) and predicting the risk of rupture is crucial for surgical management. Previously, the maximum diameter criterion was the only one used regularly to evaluate rupture risk. However, there is a trend towards using hemodynamic and biomechanical conditions, which relies on computational study. Most studies have looked at the hemodynamic and biomechanical properties at the peak growth or rupture point; however, a better understanding of the changes throughout the growth process may help clinicians in daily care to prevent SCA and sudden rupture. This approach is relevant in areas that need more resources to utilize in research and development programs. These changes are examined in the context of the current study by analyzing fluid-structure interaction (FSI), which reveals the progressive stages of aortic aneurysm formation.

As explained in [15], influenza results in about five million severe illnesses and 650,000 respiratory deaths yearly. The current escalation and influenza surveillance are crucial in adequately utilizing resources, cost-effectiveness and avoiding deaths. This work brings a new data-driven method for local early detection and prediction of influenza epidemics based on diagnostic data from over 3,000 clinics in Malaysia. This diagnostic dataset's

DOI: https://doi.org/10.54216/JAIM.080204
Received: March 28, 2024 Revised: June 05, 2024 Accepted: November 08 2024

reliability and speedy availability make the methodology more effective. A new RI was created from this information, and by performing statistical analysis of weekly RI outliers, one can detect regions likely to experience an outbreak; Furthermore, an ensemble learning model was used for outbreak prediction, and cross-validation was performed to tune the model's parameters. Evaluation proceeded from testing data that also eliminated biases to ascertain the model's high performance, with an average of 75% in recall, 74% in precision, and 83% inaccuracy in the five regions. The trends were also validated using Google flu trends, news reports and surveillance data from the World Health Organization.

In the article marked as [16], influenza, which is still a significant public health problem, contributes to the incidence of severe illness of about five million and respiratory deaths of 650,000 per year. Meaningful prediction of flu incidences plays a role in facilitating timely resource mobilization, minimization of expenditure and saving lives. This work presents a methodology that provides the real-time analysis and prediction of influenza-like illness incidence in a small geographic area based on diagnostic data from over three thousand clinics in Malaysia. This diagnostic dataset is dependable and easily accessible, and this process entails the creation of a new region index (RI). Perspective epidemics are detected based on the statistical evaluation of deviations of weekly RI values from the trend. In addition, an ensemble learning model is employed to predict the probability of an outbreak, and for further fine-tuning of its hyperparameters, a cross-validation test is done. The evaluation data were used for an accurate test of the performance of each model employed, with the average test results as follows: 75% recall, 74% precision, and 83% accuracy of test data set across five regions. These results were also confirmed using information from Google Flu Trends, newspaper articles and WHO surveillance information.

To the extent that this can be described, as discussed in [17], avian influenza has been a threat to economies and human health for quite a long time; it has always emerged and spread unpredictably. The virus can spread through human and poultry transit through wild bird movement, which complicates the spread of the virus. For this purpose, the study designed a decision-support framework to enhance the timely and adequate prediction of AI events. This system improves perspectives on the environment and provides opportunities for quick reactions for managers. In this way, the proposed framework functions as risk patterns developed from common parts deduced beforehand and combined into a fully integrated knowledge base. These authorities can go straight to this knowledge base and input queries or use various in-built analyses to compute future risks at various geographical tiers. The assessment of the performance of the proposed system revealed an average sensitivity of 69.70% and specificity of 85.50%, thus, the potential to aid the healthcare authorities in addressing outbreak concerns in advance.

Based on the work in [18], influenza impacts 3 to 5 million people annually, leading to 290,000 to 650,000 deaths worldwide. Numerous countries have implemented influenza surveillance systems to mitigate these fatalities and gather early warning data. However, these systems often need 1 to 2 weeks delays between the actual onset of outbreaks and the release of surveillance data, limiting their effectiveness. To overcome this issue, novel surveillance and prediction methods leveraging real-time internet data, such as search queries, social media posts, and news articles, have been explored. Current approaches typically involve extracting online data and using machine learning to predict influenza outbreaks in a classification framework. Despite their promise, many of these methods rely on subjectively selected training data, making it challenging to accurately capture the underlying patterns in the data. There is, therefore, a pressing need for new techniques that focus on extracting training data that better reflects the latent characteristics of the information, thereby enhancing the predictive accuracy of these systems.

As summarized in the paper [19], this notion is evidenced by the H1N1 pandemic in 2009 that led to approximately 203,000 global fatalities. The means for the spatiotemporal dynamics of influenza during its incubation period should be accurately predicted in order to prevent the consequences of potential pandemic minimization. To conclude, this study proposes a novel prediction system for ILI that considers latent temporal and spatial dependencies. The system employs a multi-step approach: First, it applies Gaussian function models and multivariate polynomial regression to study the temporal-spatial pattern of ILI data; second, delay-coordinate embedding is employed to reconstruct the phase space and investigate the dynamic feature of a 1-D time series of ILI; and finally, a DRBFNN, which connects an observed data space with a reconstructed phase space, is the core of the prediction method in the current study. The assessment of the system's performance proves that the regression models integrated with spatial distribution information can tame missing data concerns, and the DRBFNN successfully models the trends of ILI for the subsequent year. Furthermore, the system integrates a model-free control method; no fixed equations define the inputs and outputs and can be applied to other fields, such as meteorology, industry, and finance. For instance, one application was used to predict the Standard & Poor's 500 index and inform the tendency of open prices for the next eight trading days.

In the study cited [in the text as [20], for example, avian influenza (AI) is discussed as a many-faceted and as of yet not fully understood illness, especially concerning such issues as, for instance, its sources, concurrent infections, and other environmental parameters. HPAIVs are known to exist and have been described in great detail. However, studies on LPAIV that cause mild conditions in chickens are fewer, partly because of FI and

DOI: https://doi.org/10.54216/JAIM.080204
Received: March 28, 2024 Revised: June 05, 2024 Accepted: November 08 2024

because they provide little evidence about AI ecology and host immune responses. Interactions between LPAIVs and HPAIVs are expected within their respective ecological niches, given that both subtypes exist simultaneously in specified geographic regions. This work follows an international perspective to develop and forecast the distribution of ecological adaptability of LPAI across the Pacific Rim using machine learning approaches on open-source data and geographic information system (GIS) on a 5 km pixel resolution for accurate estimations. The data consists of approximately 40,827 lab-analyzed field records from Japan, Russia, Vietnam, Mongolia, Alaska and the Influenza Research Database (IRD) US Department of Agriculture (USDA). Host sampling incorporated 157 hosts and 110 LPAIVs in thirty-two species and identified significant hosts as Muscovy ducks, Mallards, Whistling Swans, and gulls related to industrialization effects on human-wildlife interactions. The findings of this work show the effectiveness of extensive data mining, ecological modeling, and machine learning approaches for discovering AI sources and forecasting epidemics. Further research must follow suit in the study of HPAI and similar threats.

As stated in [21], one of the most significant challenges in combating influenza A viruses is their relatively high mutation rate. To this end, analytical and machine-learning models have been considered to enhance antigenicity prediction. This work applied a deep CNN termed DL-FIA for the first time, systematically evaluating 566 amino acid characteristics and 141 amino acid substitution matrices for their predictive potential. The proposed model structure was further optimized by particle swarm optimization, and the neural network generated a blind validation accuracy of 95.8% to set the work for the existing models. The CNN model was also used to compare recommended vaccines from 1997 to 2011, along with traditional experimental approaches and WHO recommendations. The study showed that the CNN model provided higher accuracy than the WHO lost method, which usually adopted a virus strain with minor changes from one year to another, implying slow changes when coverage was reduced. Unlike the strains identified from the bacterial culture method, the strains selected by the CNN model were less consistent year by year. However, the CNN model had more constant coverage, indicating more flexibility or less changeable from the virus.

As pointed out [22], influenza viruses remain a menace to population health and cause seasonal epidemics and occasional pandemics. The frequent mutation of these viruses negatively impacts antiviral chemoprevention; thus, detection and accurate prediction of the virulence of influenza strains are essential for effective influenza monitoring and control. This work involves developing a novel weighted ensemble CNN model called VirPreNet to predict the virulence of Influenza A viruses based on all eight genomic segments. To determine the difference between avirulent and virulent infections, the model employs the mouse lethal dose fifty. This approach uses a numerical encoding system called ProtVec to transform the amino acids into a protein and across the eight segments for distributed processing of biological sequences. The final ensemble CNN is trained using influenza datasets of each segment and has been employed as the core of VirPreNet. The base model predictions are combined and calibrated through the last fully connected layer to achieve the final estimation. Results prove that VirPreNet provides improved results compared to the baseline methods and yields high performance on independent testing data sets when integrated with the proposed architecture and ProtVec.

In the research discussed in [23], the problem of diagnostics and prevention of such infectious diseases as influenza and Ebola is described as one of the most important but, at the same time, very complex tasks, which need a proper characterization of disease dynamics and epidemic processes. Although computational epidemiology can simulate the disease transmission process and the contact network, it is diffusion in processing actual-time, high-resolution surveillance data. In contrast, surveillance facts are timely and detailed in social media but are not connected to disease models and contact networks. To address these gaps, the study introduces a semi-supervised neural network approach that will integrate computational epidemiology and social media mining for influenza epidemiological modeling. Based on the disease model contact network, this framework can learn health states and intervention actions from social media users in real time to make a more accurate and efficient disease diffusion model. An online optimization algorithm updates the model through iteration to enable this interactivity. The experimental results show that the proposed approach is more accurate in outbreak prediction than the conventional methods in epidemic modeling as it depicts the clinical longitudinal severity and dispersion as well as cross-sectional morbidity and mortality by incorporating individual characteristics that play a critical role in disease transmission and progression.

As mentioned in [24], early diagnosis of COVID-19 cases is essential to minimize its transmissibility and promptly provide medical intervention to those infected. The study presents a COVID-19 diagnosis and prediction model known as AIMDP, an automated model whose purpose is to detect cases of COVID-19 from viral pneumonia using CT chest images. It uses convolutional neural networks (CNNs) to process many CT images to quickly and accurately predict COVID-19 cases and help contain the disease. The whale optimization algorithm (WOA) selected the most suitable clinical signs for diagnosis. The confirmed AIMDP's performance during a set of experiments, considering its results according to the area under the receiver operating characteristic (AUC-ROC) curve, the positive predictive value (PPV), the negative predictive rate (NPR), as well as the negative predictive value (NPV). In order to evaluate the AIMDP model, it was applied to a set of hundreds of real-world cases and

DOI: https://doi.org/10.54216/JAIM.080204
Received: March 28, 2024 Revised: June 05, 2024 Accepted: November 08 2024

CT images; the models showed a 96% AUC and 98% overall accuracy in COVID-19 diagnosis and outcompeted other modern diagnostic and prediction models. These findings have implications that AIMDP is a promising tool to be used in the early identification of cases of COVID-19.

Based on the research conducted in [25], it has been identified that early identification of the influenza virus is appealing due to the high mortality rates related to influenza diseases. Through the research, a machine-learning technique for the detection of the influenza virus using images from an influenza detection kit is proposed. The proposed model uses convolutional neural networks (CNNs), an image classification model known for its efficiency in this task due to the specificity of the model used in the study, which was constructed and optimized for use. An architecture search algorithm, Bayesian optimization, and hyperband (BOHB) allowed for adequately selecting the CNNs hyperparameters, resulting in an efficient classification system. The developed 2D CNN model achieved an overall accuracy of 90.14%, which means the potential to become a reliable diagnostic instrument for identifying the influenza virus from the standard samples using the images from the detection kits.

As stated in [26], vaccinations are the most effective and inexpensive way of preventing and controlling the flu, with surface antigen – HA being the main target of the nelizing antibodies. On the other hand, antigenic drift due to constant changes in the HA sequence may result in strain/ Vaccine mismatch. Forecasting possible antigenic variants and stratifying viruses is necessary to choose appropriate strains for the vaccine. To fill the gap, the study proposes PREDAC-CNN, a novel CNN model for tracking the antigenic evolution of seasonal influenza a viruses. Using the spatial feature extraction ability of CNNs for the HA1 sequence, we examine the interactions between the specific amino acid sites. PREDAC-CNN also removes the extra noise of non-critical amino acid embedding's from the physicochemical properties that determine antigenicity, which is essential. It faithfully captures the impacts of point mutations on antigenic evolution and identifies the dominant antigenic lineages, especially for A/H3N2 (1968–2023) and A/H1N1 (1977–2023) viruses. This is mainly because the model under study was mapped through 5-fold cross-validation and retrospective testing to offer improved accuracy over other existing approaches, thereby making it play a central role in enhancing the accuracy of the vaccine-recommended strain. PREDAC-CNN is accessible at http://predac-cnn.cloudna.cn, offering the spatial optimized framework for computing the antigenicity of those human influenza viruses.

As described in [27], NS statistical surveillance systems and web data analysis methods have been applied to forecast Influenza epidemics. However, most existing models are adapted to make a short-term forecast for a week at most. However, accurate early warning and early intervention to prevent influenza epidemics require forecasts from anywhere from two to ten weeks. The work presents an idea using the time-precedence identified between specific web data and subsequent flu occurrences. When using web search queries, the presence of such words as 'colds' in the prior weeks of an influenza pandemic is the most dominant. Using this temporal relationship, the researchers constructed a long-term forecast of influenza, which incorporated early web data to enhance the system's accuracy. Experiments covered the choice of relevant web data, the analysis of regional dependence of the model and the comparison of the predictability accuracy over different time intervals. The proposed model has discovered a correlation of 0.87 for up to TEN weeks to early warning systems, outperforming other current mechanisms and, therefore, has a better solution for long-term influenza outbreak management.

As shown in the study conducted in [28], HPAI, particularly H5N1, H5N8, and H5N6 epidemics, have challenged poultry production in South Korea from 2003 to 2017. The last outbreaks of these HPAI subtypes have caused widespread concern, specifically in the country's southern region, which accounts for 58.3% of the cases. This raised the need to study spatial risk factors for extended periods and predict future HPAI cases. To overcome this, spatial descriptors of twelve variables that characterize HPAI-infected premises (IPs) were studied across 88 H5N1, 339 H5N8, and 335 H5N6 cases. They constructed two Bayesian logistic regression models from their case-control study of the H5N8 epidemic and a machine learning model known as extreme gradient boosting, or XGBoost. These models were later applied to the predictability of the risk of H5N1 and H5N6 outbreaks. This study showed that the interface between domestic ducks and live bird markets increased the incidences of HPAI risks. In general, the two predictive models were highly effective; the specificity of the Bayesian model was above 0.82, and the XGBoost model was above 0.97, pointing to spatial characteristics of interest in averting HPAI risk. These results help design effective prevention and control interventions that reduce the consequences and effects of HPAI outbreaks across the poultry production system.

As outlined in the paper [29], there exists an intrinsic difficulty in Developing predictive capabilities for the future state evolution of nonlinear dynamical systems; coupling these predictions increases the degree of difficulty when attempting to address problems where the dynamics are chaotic, Black Swan, or where there exists a shift in the base processes governing the dynamics of the system. The long history of prediction methods mainly depends on the sequences of previous observations and presupposes statistical stationarity, which could be more efficient in such cases. To address these limitations, the study introduces a comprehensive methodology comprising: The CDE mobile can simultaneously develop the following functional modules: (1) serve as a global and local prediction algorithm to deal with the range of complex systems; (2) switch between predictions between global and local

DOI: https://doi.org/10.54216/JAIM.080204
Received: March 28, 2024 Revised: June 05, 2024 Accepted: November 08 2024

algorithms; and (3) track hypothetical predictions based on the untouched systems using these predictions when the system returns to untouched settings. This new approach is based on Koopman operator theory, which provides a model-free, data-driven approach that effectively accommodates dynamic variations. Despite the guidance given via the COVID-19 and influenza cases, the methodology can be applied across multiple domains beyond epidemiology and is a solid framework for forecasting in complex, constantly changing environments.

As we mentioned in the paper [30], it is difficult to detect early warning signs of viral outbreaks, like the ILI, since the rate of viral expansion is exponential. However, early diagnosis forms the basis of any public health intervention. The study has prospectively suggested a new approach that applies the centrality of nodes in social media networks to gauge early signals of ILI outbreaks. In this study, using a massive dataset of tweets collected from Twitter over three years, it is evidenced that highly central users with high out-degrees can help act as social network sensors for early detectable signs of ILI outbreaks. Unlike the 'sensors,' these do not presuppose observation of the overall population, and therefore, the approach is plausible, more effective and sensitive to privacy. Also, the study examines the behavioral and content-based characteristics: while effective-sensing users are involved in discussions about local news, language, politics or government, the project identifies them as conceptually different from typical users. Not only does this method discover a smaller and more accurate number of social sensors, but it also increases operational effectiveness without requiring much data retrieval, making it a more helpful and less intrusive solution for the early identification of emerging viral conditions.

Amid flu, as explained in the paper [31], regularly evolves to adapt to human immunity to ensure the vaccine's functionality, the antigenic differences among strains should be monitored. Differences such as these can be determined through conventional serological techniques, which are usually lengthy and require a large workforce to accomplish; hence, there is a demand for more computational approaches. This work proposes MetaFluAD as a novel meta-learning framework for quantitatively forecasting antigenic distances of different Influenza strains. MetaFluAD combines the strains represented within the weighted attributed network using a strategy similar to antigenic cartography. The method also includes an encoder based on GNN and a meta-learning framework to obtain a united space combining antigenically and genetically defined characteristics. This allows knowledge to be passed from one influenza subtype to another, improving performance even with little data. As presented, MetaFluAD has provided stable and reproducible outcomes for various subtypes, including A/H3N2, A/H1N1, A/H5N1, B/Victoria, and B/Yamagata. Moreover, MetaFluAD, based on the integration of GNN-based encoding and meta-learning algorithm, presents a vital perspective for accurate prognosis of antigenic distances, an identification of the significant clusters of antigenic variations among the seasonal IAVs to strengthen the creation of efficient vaccines and enhancing the viral surveillance systems.

According to the publication described [32], influenza is a global health issue that results in a considerable burden every other year, thus informing a need for accurate models to facilitate the preparation process for hospitals, pharmaceutical companies, and governments. However, seasonality and the sporadic flu make forecasting challenging, given the many uncertainties surrounding influenza. The forecasting methods are tied to current and past data within user-specified temporal windows and do not consider conditions outside those windows. To address these issues, we present the Dynamic Virtual Graph Significance Networks (DVGSN), a graph-based algorithm that can dynamically and supervise learn at all-time steps and is not confined to time-windows inference from similar "infection situations." This approach also allows representation learning on the virtual graph to be dynamic and adapt to alterations such as seasonal fluctuation with the added benefit of pandemic consideration without much knowledge in that field. Real-world Influenza datasets show that DVGSN has comparable results with state-of-the-art methods that manifold superior to the approaches of another group of algorithms. This is the first time a dynamic virtual graph has been used in a supervised learning context for time series prediction. Also, the specific method under discussion provides relatively high interpretability, which will increase its applicability in public health and life sciences.

According to the research in the article [33], PPSIV is crucial for preventing many deaths annually due to seasonal flu; however, modifications to these vaccines are required periodically due to the ongoing genetic reassortment of the influenza a virus segment. The decision to formulate the laminated distal subpopulations usually involves appraising the existing dominant strains. Regrettably, the time it takes to manufacture and distribute vaccines increases the likelihood of developing new variants that could reduce the efficiency of vaccines. Thus, understanding the dynamics of its evolution can prove very valuable in evaluating and selecting vaccines. In response, we present FluPMT – a sequence prediction model based on an encoder-decoder architecture that predicts the HA protein sequence of the next season's dominant strain using evolutionary patterns of Influenza A viruses. In detail, we use the temporal data to capture the above evolution trends and attention structures to model dependencies between the sequences' amino acids. Moreover, we incorporate antigenic distance prediction by employing graph network representation learning into a model as a secondary task in the multi-task learning approach. Based on experimental evaluations of two influenza datasets, FluPMT shows its outstanding prediction

DOI: https://doi.org/10.54216/JAIM.080204
Received: March 28, 2024 Revised: June 05, 2024 Accepted: November 08 2024

capability and will be beneficial in understanding the evolution of the virus and guiding the evaluation and production of vaccines.

Table 1 overviews the most relevant papers on modeling and forecasting influenza epidemics using machine learning. Each citation includes the article's scope, the techniques used in the research question, and the main conclusions reached with accompanying evidence to demonstrate the discipline's range and progress. Some of these studies considered purely statistical studies, built machine learning models, or used creative data collection methods, all of which prove the effectiveness of the techniques in tackling public issues revolving around the outbreak of influenza and other viruses. Thus, resuming the information in the table is intended to explain the main outlines of research done already and point to the areas for possible research in the future.

Table 1: Summary of Literature Review

Reference	Area of Focus	Methodology	Key Findings
[14]	Long-term influenza outbreak forecasting	Correlation of web data with time precedence	Achieved high correlation (0.87) for long-term flu forecasts using web data.
[15]	Spatial risk factor prediction for HPAI outbreaks	Bayesian logistic regression models, XGBoost	High specificity (>0.82) in predicting HPAI outbreak risks based on spatial characteristics.
[16]	Methodology for complex system prediction applicable to influenza	Koopman operator theory for model-free, data-driven approach	Provided a flexible framework for forecasting in complex, dynamic environments.
[17]	Early detection of ILI outbreaks using social network analysis	Centrality analysis of social media networks	We identified highly central users as effective sensors for early detection of ILI outbreaks.
[18]	Antigenic distance prediction of influenza strains	MetaFluAD - meta-learning framework with GNN-based encoding	Improved prediction of antigenic distances for various influenza subtypes.
[19]	Influenza forecasting using dynamic virtual graph networks	Graph-based algorithm for supervised learning	Comparable results with state-of-the-art methods; high interpretability.
[20]	Predicting HA protein sequence of dominant influenza strains	FluPMT - sequence prediction model with multi-task learning	Enhanced understanding of influenza evolution and vaccine strain evaluation.
[21]	Antigenicity prediction and vaccine recommendation for H3N2	Deep CNN model (DL-FIA) with optimization	We achieved high blind validation accuracy (95.8%) for antigenicity prediction.
[22]	Virulence prediction of Influenza A viruses	Weighted ensemble CNN model (VirPreNet)	High performance in predicting virulence using genomic segment data.
[23]	Integrating social media mining with epidemiology	Semi-supervised neural network approach	Enhanced outbreak prediction by combining social media data with disease models.
[24]	Early diagnosis of COVID-19 using CT images	AIMDP - CNN-based diagnostic model	Achieved 96% AUC and 98% overall accuracy in COVID-19 diagnosis.
[25]	Influenza virus detection using image classification	2D CNN model with hyperparameter optimization	She demonstrated 90.14% overall accuracy in detecting influenza virus from kit images.

DOI: https://doi.org/10.54216/JAIM.080204
Received: March 28, 2024 Revised: June 05, 2024 Accepted: November 08 2024

[26]	Tracking antigenic evolution of influenza, A viruses	PREDAC-CNN using spatial feature extraction	Improved accuracy in predicting antigenic evolution, aiding vaccine development.
[27]	Long-term influenza outbreak prediction	Time-precedence correlation of web data	Correlation of 0.87 for early warning systems, outperforming existing mechanisms.
[28]	Spatial risk factor analysis for HPAI outbreaks in poultry	Bayesian logistic regression, XGBoost	High specificity (>0.82) and predictive accuracy based on spatial risk factors.
[29]	Complex system forecasting applicable to influenza	Koopman operator-based prediction algorithm	Flexible and adaptable for dynamic environments, effective for forecasting.
[30]	Early detection of ILI via social network analysis	Centrality analysis of social media data	Effective early warning system using social media sensors.
[31]	Antigenic distance prediction of influenza viruses	MetaFluAD with GNN-based encoding and meta-learning	Improved performance in predicting antigenic distances, aiding vaccine design.
[32]	Influenza Prediction with pandemic-awareness	Dynamic Virtual Graph Significance Networks (DVGSN)	Effective in adapting to changing conditions and providing high interpretability.
[33]	Predicting the evolution of dominant influenza strains	FluPMT - multi-task learning model	It has enhanced the understanding and prediction of virus evolution and vaccine evaluation.

As a synthesis of the reviewed literature, it is evident that considerable strides have been made concerning prediction research on influenza. Different techniques have been utilized, including but not limited to statistical modeling, machine learning, and data mining and soliciting information from social networks, and each has its merits and demerits. Although notable achievements have been made, there is room for enhancing the accuracy, relevance, and usability of prediction models. In this context, it is also essential to specify two additional directions for future research: the evolution of more straightforward and more flexible models for such processes as the dynamics of the influenza virus and the dynamics of human society. Overall, it is worth continuing the work on influenza prediction research, as doing so helps to enhance the mapping of this pandemic disease and its consequences.

3. Conclusion

The conclusion is therefore justified, and the main argument stated is that incorporating machine learning approaches for the forecast or detection of influenza can significantly improve society's response to this endemic. Recent innovations in data analytics, especially the integrated data types, have shown great promise in the effectiveness and efficiency of predicting and forecasting outbreaks, especially epidemics. With the support of all the technologies, such as social media, search engines, and clinical reports, interesting new data - inaccessible only a few years ago - can be generated by applying advanced computation models that employ algorithms to analyze information. While barriers remain to be overcome, such as the need for quality-controlled data and understanding how models make predictions, the successful deployment of machine learning in practice, particularly in public health, hinges on overcoming these challenges. This will require sustained collaboration between data scientists, epidemiologists, and healthcare providers in these models' evolution and scaling up. Future studies will aim to sequence genome variation and how different virus clades interact with model organisms. A more detailed focus on hybrid approaches that juxtapose classical epidemiological models and machine learning algorithms prompted the enhancement of forecasting models. Furthermore, sentiments and their communication channels can help understand influenza transmission patterns and their intensity. Ultimately, advancing and developing ideas on potential machine-learning techniques for influenza forecasting is prudent. Health professionals are already

DOI: https://doi.org/10.54216/JAIM.080204
Received: March 28, 2024 Revised: June 05, 2024 Accepted: November 08 2024

involved in activities related to forecasting, and these efforts could be significantly enhanced if management decides to invest in such measures.

Funding: "This research received no external funding"
Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] Q. Chen et al., "Prediction of influenza outbreaks in Fuzhou, China: comparative analysis of forecasting models," BMC Public Health, vol. 24, no. 1, pp. 1–12, Dec. 2024, doi: 10.1186/S12889-024-18583-X/FIGURES/5.

[2] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings, 2013.

[3] S. A. Ajagbe and M. O. Adigun, "Deep learning techniques for detection and prediction of pandemic diseases: a systematic literature review," Multimed Tools Appl, vol. 83, no. 2, pp. 5893–5927, Jan. 2024, doi: 10.1007/S11042-023-15805-Z/TABLES/3.

[4] S. Wei et al., "The prediction of influenza-like illness using national influenza surveillance data and Baidu query data," BMC Public Health, vol. 24, no. 1, pp. 1–12, Dec. 2024, doi: 10.1186/S12889-024-17978-0/FIGURES/3.

[5] T. J. Kieran, X. Sun, T. R. Maines, and J. A. Belser, "Machine learning approaches for influenza A virus risk assessment identifies predictive correlates using ferret model in vivo data Check for updates", doi: 10.1038/s42003-024-06629-0.

[6] P. Mahajan, S. Uddin, F. Hajati, M. A. Moni, and E. Gide, "A comparative evaluation of machine learning ensemble approaches for disease prediction using multiple datasets," Health Technol (Berl), vol. 14, no. 3, pp. 597–613, May 2024, doi: 10.1007/S12553-024-00835-W/TABLES/16.

[7] A. Hassan Zadeh, H. M. Zolbanin, R. Sharda, and D. Delen, "Social Media for Nowcasting Flu Activity: Spatio-Temporal Big Data Analysis," Information Systems Frontiers, vol. 21, no. 4, pp. 743–760, Aug. 2019, doi: 10.1007/S10796-018-9893-0/TABLES/9.

[8] F. Liang, P. Guan, W. Wu, and D. Huang, "Forecasting influenza epidemics by integrating internet search queries and traditional surveillance data with the support vector machine regression model in Liaoning, from 2011 to 2015," PeerJ, vol. 2018, no. 6, p. e5134, Jun. 2018, doi: 10.7717/PEERJ.5134/SUPP-4.

[9] H. Bandi and D. Bertsimas, "Optimizing Influenza Vaccine Composition: From Predictions to Prescriptions," Sep. 18, 2020, PMLR. Accessed: Oct. 15, 2024. [Online]. Available: https://proceedings.mlr.press/v126/bandi20a.html

[10] X. Li, Y. Li, X. Shang, and H. Kong, "A sequence-based machine learning model for predicting antigenic distance for H3N2 influenza virus," Front Microbiol, vol. 15, p. 1345794, Jan. 2024, doi: 10.3389/FMICB.2024.1345794/BIBTEX.

[11] J. Henriques, T. Rocha, P. de Carvalho, C. Silva, and S. Paredes, "Interpretability and Explainability of Machine Learning Models: Achievements and Challenges," IFMBE Proc, vol. 108, pp. 81–94, 2024, doi: 10.1007/978-3-031-59216-4_9/TABLES/1.

[12] J. Li et al., "Machine Learning Methods for Predicting Human-Adaptive Influenza A Viruses Based on Viral Nucleotide Compositions," Mol Biol Evol, vol. 37, no. 4, pp. 1224–1236, Apr. 2020, doi: 10.1093/MOLBEV/MSZ276.

[13] R. Agarwal et al., "Neural Additive Models: Interpretable Machine Learning with Neural Nets," Adv Neural Inf Process Syst, vol. 6, pp. 4699–4711, 2021.

[14] A. Zan et al., "DeepFlu: a deep learning approach for forecasting symptomatic influenza A infection based on pre-exposure gene expression," Comput Methods Programs Biomed, vol. 213, p. 106495, Jan. 2022, doi: 10.1016/J.CMPB.2021.106495.

[15] L. Du and Y. Pang, "A novel data-driven methodology for influenza outbreak detection and prediction," Scientific Reports 2021 11:1, vol. 11, no. 1, pp. 1–16, Jun. 2021, doi: 10.1038/s41598-021-92484-6.

[16] L. Du and Y. Pang, "A novel data-driven methodology for influenza outbreak detection and prediction," Scientific Reports 2021 11:1, vol. 11, no. 1, pp. 1–16, Jun. 2021, doi: 10.1038/s41598-021-92484-6.

[17] S. Yousefinaghani, R. A. Dara, Z. Poljak, and S. Sharif, "A decision support framework for prediction of avian influenza," Scientific Reports 2020 10:1, vol. 10, no. 1, pp. 1–14, Nov. 2020, doi: 10.1038/s41598-020-75889-7.

DOI: https://doi.org/10.54216/JAIM.080204
Received: March 28, 2024 Revised: June 05, 2024 Accepted: November 08 2024

[18] B. Jang, I. Kim, and J. W. Kim, "Effective Training Data Extraction Method to Improve Influenza Outbreak Prediction from Online News Articles: Deep Learning Model Study," JMIR Med Inform 2021;9(5):e23305 https://medinform.jmir.org/2021/5/e23305, vol. 9, no. 5, p. e23305, May 2021, doi: 10.2196/23305.

[19] X. Guo, N. N. Xiong, H. Wang, and J. Ren, "Design and Analysis of a Prediction System about Influenza-Like Illness from the Latent Temporal and Spatial Information," IEEE Trans Syst Man Cybern Syst, vol. 52, no. 1, pp. 66–77, Jan. 2022, doi: 10.1109/TSMC.2020.3048946.

[20] M. Gulyaeva et al., "Data mining and model-predicting a global disease reservoir for low-pathogenic Avian Influenza (AI) in the wider pacific rim using big data sets," Scientific Reports 2020 10:1, vol. 10, no. 1, pp. 1–11, Oct. 2020, doi: 10.1038/s41598-020-73664-2.

[21] E. K. Lee, H. Tian, and H. I. Nakaya, "Antigenicity prediction and vaccine recommendation of human influenza virus A (H3N2) using convolutional neural networks," Hum Vaccin Immunother, vol. 16, no. 11, pp. 2690–2708, Nov. 2020, doi: 10.1080/21645515.2020.1734397.

[22] R. Yin, Z. Luo, P. Zhuang, Z. Lin, and C. K. Kwoh, "VirPreNet: a weighted ensemble convolutional neural network for the virulence prediction of influenza A virus using all eight segments," Bioinformatics, vol. 37, no. 6, pp. 737–743, Mar. 2021, doi: 10.1093/BIOINFORMATICS/BTAA901.

[23] L. Zhao et al., "Online flu epidemiological deep modeling on disease contact network," Geoinformatics, vol. 24, no. 2, pp. 443–475, Apr. 2020, doi: 10.1007/S10707-019-00376-9/FIGURES/16.

[24] A. Ella Hassanien, V. Snasel, S. M. Elghamrawy, A. Ella Hassnien, and C. Author, "Optimized Deep Learning-Inspired Model for the Diagnosis and Prediction of COVID-19", doi: 10.32604/cmc.2021.014767.

[25] J. Lee et al., "End-to-end Convolutional Neural Network Design for Automatic Detection of Influenza Virus," IEEE Transactions on Smart Processing & Computing, vol. 10, no. 1, pp. 31–36, Feb. 2021, doi: 10.5573/IEIESPC.2021.10.1.031.

[26] J. Meng et al., "PREDAC-CNN: predicting antigenic clusters of seasonal influenza A viruses with convolutional neural network," Brief Bioinform, vol. 25, no. 2, pp. 1–12, Jan. 2024, doi: 10.1093/BIB/BBAE033.

[27] B. Jang, I. Kim, and J. W. Kim, "Long-Term Influenza Outbreak Forecast Using Time-Precedence Correlation of Web Data," IEEE Trans Neural Netw Learn Syst, vol. 34, no. 5, pp. 2400–2412, May 2023, doi: 10.1109/TNNLS.2021.3106637.

[28] D. S. Yoo, B. C. Chun, K. Hong, and J. Kim, "Risk Prediction of Three Different Subtypes of Highly Pathogenic Avian Influenza Outbreaks in Poultry Farms: Based on Spatial Characteristics of Infected Premises in South Korea," Front Vet Sci, vol. 9, p. 897763, May 2022, doi: 10.3389/FVETS.2022.897763/BIBTEX.

[29] I. Mezić et al., "A Koopman operator-based prediction algorithm and its application to COVID-19 pandemic and influenza cases," Scientific Reports 2024 14:1, vol. 14, no. 1, pp. 1–13, Mar. 2024, doi: 10.1038/s41598-024-55798-9.

[30] D. Martín-Corral, M. García-Herranz, M. Cebrian, and E. Moro, "Social media sensors as early signals of influenza outbreaks at scale," doi: 10.1140/epjds/s13688-024-00474-1.

[31] Q. Jia, Y. Xia, F. Dong, and W. Li, "MetaFluAD: meta-learning for predicting antigenic distances among influenza viruses," Brief Bioinform, vol. 25, no. 5, p. 395, Jul. 2024, doi: 10.1093/BIB/BBAE395.

[32] J. Zhang, P. Zhou, Y. Zheng, and H. Wu, "Predicting influenza with pandemic-awareness via Dynamic Virtual Graph Significance Networks," Comput Biol Med, vol. 158, p. 106807, May 2023, doi: 10.1016/J.COMPBIOMED.2023.106807.

[33] C. Cai, J. Li, Y. Xia, and W. Li, "FluPMT: Prediction of Predominant Strains of Influenza A Viruses Via Multi-Task Learning," IEEE/ACM Trans Comput Biol Bioinform, 2024, doi: 10.1109/TCBB.2024.3378468.

DOI: https://doi.org/10.54216/JAIM.080204
Received: March 28, 2024 Revised: June 05, 2024 Accepted: November 08 2024