Journal of Artificial Intelligence and Metaheuristic (JAIM)

♦ ISSN: 2833-5597 Vol: 08, No. 02, 2024 ♦ PP. 20–26

Article DOI: https://doi.org/10.54216/JAIM.080204



REVIEW ARTICLE

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: Ehsan.k@uark.edu

Received: March 28, 2024 Revised: June 05, 2024 Accepted: November 08, 2024 ⋆ Corresponding author

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 are not limited to data issues, model explainability 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].

Complementing traditional surveillance with auxiliary, heterogeneous data is one approach to improving influenza prediction; a deep-learning model using multisource data has been developed and validated for a megacity [4].

In recent years, there has been increasing attention to machine-learning methods for influenza prediction. Neural-network forecasting models can use web-search activity and quantify predictive uncertainty, illustrating how ML can capture patterns that conventional statistical methods may overlook [5]. Ensemble forecasting methods can further combine

© 2024 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

model outputs to improve influenza predictions across different settings [6].

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].

Sequence-based machine-learning models have also been used to estimate antigenic distances between influenza A strains from their attributes and network relationships, supporting the monitoring of antigenic evolution [10].

Model explainability remains a major challenge: predictions from complex machine-learning models can be difficult to interpret, motivating methods for transparent and accountable AI [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 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% accuracy 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 eas-

ily 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 the 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 inbuilt 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' delay 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 as [20], for example, avian influenza (AI) is discussed as a many-faceted and as yet not fully understood illness, especially concerning such issues as 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 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 consist 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 list 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 changeability from the virus.

As pointed out in [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 difficult to process 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 AIMDP's performance was confirmed 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 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.

Vaccination is an important means of preventing influenza, but antigenic drift in hemagglutinin (HA) can lead to vaccine mismatch. Liu et al. [26] developed PREDAC-H1pdm to characterize antigenic evolution among A/(H1N1)pdm09 viruses. By analyzing HA sequence features and antigenic relationships, the model identifies variants and clusters associated with evolving antigenic patterns, providing information relevant to influenza surveillance and vaccine-strain selection.

As described in [27], 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. 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 for 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.

Rodríguez et al. [29] introduced epidemiologically informed neural networks (EINNs) that learn latent epidemic dynamics and transfer the learned information to data-driven neural models. Evaluated on influenza and COVID-19 forecasting across U.S. states and Health and Human Services regions, the framework improved short- and long-term forecasting over the compared approaches.

Athanasiou et al. [30] developed and validated long short-term memory models for influenza-like illness forecasting using surveillance, weather, and Twitter data. Their comparisons examine the contribution of each source and the value of combining them for short-term predictions, demonstrating how online signals can complement conventional epidemiological observations.

Yin et al. [31] developed CLCAP, a contrastive-learning framework based on convolutional neural networks, to predict influenza A antigenicity. Evaluations across H1N1, H3N2, and H5N1 show improved prediction over the compared baselines, suggesting that contrastive learning can help extract informative antigenic representations for surveillance and vaccine research.

According to the publication described in [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, the Dynamic Vir-

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.
[26]	Antigenic evolution of A/(H1N1)pdm09	PREDAC-H1pdm sequence-based model	Characterized antigenic patterns relevant to surveillance and vaccine-strain selection.
[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]	Influenza and COVID-19 forecasting	Epidemiologically informed neural networks (EINNs)	Improved short- and long-term forecasts across U.S. regions.
[30]	ILI forecasting from multisource data	LSTM using surveillance, weather, and Twitter data	Combined data sources support short-term ILI forecasting.
[31]	Influenza A antigenicity prediction	Contrastive-learning CNN (CLCAP)	Improved prediction across tested influenza A subtypes.
[32]	Influenza prediction with pandemic awareness	Dynamic Virtual Graph Significance Networks (DVGSN)	Effective in adapting to changing conditions and providing high interpretability.
[33]	Forecasting mutations in influenza hemagglutinin	Anomaly detection on sequential HA data	Identified mutation patterns relevant to antigenic-evolution monitoring.

tual Graph Significance Networks (DVGSN) is presented, a graph-based algorithm that can dynamically and supervisedly learn at all time steps and is not confined to time-window 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 are 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.

Garjani et al. [33] address the difficulty of identifying emerging influenza mutations against a large background of conserved hemagglutinin sequences. Their anomaly-detection approach analyzes sequential HA data to forecast mutations in influenza A, providing a complementary signal for monitoring antigenic evolution and informing vaccine-strain assessment.

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.

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

REFERENCES

[1] Z. Du, Z. Shao, X. Zhang, R. Chen, T. Chen, Y. Bai, L. Wang, E. H. Y. Lau, and B. J. Cowling. Nowcasting and forecasting seasonal influenza epidemics—china, 2022–2023. *China CDC Weekly*, 5(49):1100–1106, 2023.

[2] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. In *International Conference on Learning Representations (ICLR), Workshop Track*, 2013.

[3] O. E. Santangelo, V. Gentile, S. Pizzo, D. Giordano, and F. Cedrone. Machine learning and prediction of infectious diseases: A systematic review. *Machine Learning and Knowledge Extraction*, 5(1):175–198, 2023.

[4] L. Yang, G. Li, J. Yang, T. Zhang, J. Du, T. Liu, X. Zhang, X. Han, W. Li, L. Ma, L. Feng, and W. Yang. Deep-learning model for influenza prediction from multisource heterogeneous data in a megacity: Model development and evaluation. *Journal of Medical Internet Research*, 25(1):e44238, 2023.

[5] M. Morris, P. Hayes, I. J. Cox, and V. Lampos. Neural network models for influenza forecasting with associated uncertainty using web search activity trends. *PLoS Computational Biology*, 19(8):e1011392, 2023.

[6] L. Chen. Application progress of ensemble forecast technology in influenza forecast based on infectious disease model. *Frontiers in Public Health*, 11:1335499, 2023.

[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*, 21(4):743–760, 2019.

[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*, 6:e5134, 2018.

[9] H. Bandi and D. Bertsimas. Optimizing influenza vaccine composition: From predictions to prescriptions. In *Proceedings of Machine Learning Research*, 2020.

[10] F. Peng, Y. L. Xia, and W. Li. Prediction of antigenic distance in influenza a using attribute network embedding. *Viruses*, 15(7):1478, 2023.

[11] W. Saeed and C. Omlin. Explainable ai (xai): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems*, 263:110273, 2023.

[12] J. Li et al. Machine learning methods for predicting human-adaptive influenza a viruses based on viral nucleotide compositions. *Molecular Biology and Evolution*, 37(4):1224–1236, 2020.

[13] R. Agarwal et al. Neural additive models: Interpretable machine learning with neural nets. In *Advances in Neural Information Processing Systems*, volume 6, pages 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. *Computer Methods and Programs in Biomedicine*, 213:106495, 2022.

[15] L. Du and Y. Pang. A novel data-driven methodology for influenza outbreak detection and prediction. *Scientific Reports*, 11(1):1–16, 2021.

[16] L. Du and Y. Pang. A novel data-driven methodology for influenza outbreak detection and prediction. *Scientific Reports*, 11(1):1–16, 2021.

[17] S. Yousefinaghani, R. A. Dara, Z. Poljak, and S. Sharif. A decision support framework for prediction of avian influenza. *Scientific Reports*, 10(1):1–14, 2020.

[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 Medical Informatics*, 9(5):e23305, 2021.

[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 Transactions on Systems, Man, and Cybernetics: Systems*, 52(1):66–77, 2022.

[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*, 10(1):1–11, 2020.

[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. *Human Vaccines & Immunotherapeutics*, 16(11):2690–2708, 2020.

[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*, 37(6):737–743, 2021.

[23] L. Zhao et al. Online flu epidemiological deep modeling on disease contact network. *Geoinformatics*, 24(2):443–475, 2020.

[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, 2021.

[25] J. Lee et al. End-to-end convolutional neural network design for automatic detection of influenza virus. *IEEE Transactions on Smart Processing and Computing*, 10(1):31–36, 2021.

[26] M. Liu, J. Liu, W. Song, Y. Peng, X. Ding, L. Deng, and T. Jiang. Development of predac-h1pdm to model the antigenic evolution of influenza a/(h1n1) pdm09 viruses. *Virologica Sinica*, 38(4):541–548, 2023.

[27] B. Jang, I. Kim, and J. W. Kim. Long-term influenza outbreak forecast using time-precedence correlation of web data. *IEEE Transactions on Neural Networks and Learning Systems*, 34(5):2400–2412, 2023.

[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. *Frontiers in Veterinary Science*, 9:897763, 2022.

[29] A. Rodríguez, J. Cui, N. Ramakrishnan, B. Adhikari, and B. A. Prakash. EINNs: Epidemiologically-informed neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, no. 12, pages 14453–14460, 2023.

[30] M. Athanasiou, G. Fragkozidis, K. Zarkogianni, and K. S. Nikita. Long short-term memory-based prediction of the spread of influenza-like illness leveraging surveillance, weather, and twitter data: Model development and validation. *Journal of Medical Internet Research*, 25(1):e42519, 2023.

[31] R. Yin, B. Ye, and J. Bian. Clcap: Contrastive learning improves antigenicity prediction for influenza a virus using convolutional neural networks. *Methods*, 220:21–28, 2023.

[32] J. Zhang, P. Zhou, Y. Zheng, and H. Wu. Predicting influenza with pandemic-awareness via dynamic virtual graph significance networks. *Computers in Biology and Medicine*, 158:106807, 2023.

[33] A. Garjani, A. M. Chegini, M. Salehi, A. Tabibzadeh, P. Yousefi, M. H. Razizadeh, M. Esghaei, M. Esghaei, and M. H. Rohban. Forecasting influenza hemagglutinin mutations through the lens of anomaly detection. *Scientific Reports*, 13(1):14944, 2023.