Data-Driven Public Health Surveillance: Identifying Influenza Trends via Social

Rajeev Kumar Bhaskar[1], Balasubramaniam Kumaraswamy[2]
[1]*Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India*
[2]*Research Scholar, Department of CS & IT, Kalinga University, Raipur, India.*

KEYWORDS

Deep leaning (DL), Social media, public health, influenza, Runge Kutta Optimized Dynamic Gated Recurrent Unit (RKO-DGRU)

ABSTRACT

The internet era makes it appear outdated to monitor and identify influenza using traditional methods. Among the long-term public health problems that influenza might exacerbate include diabetes, asthma, congestive heart failure, sinus infections, ear infections, and bacterial pneumonia. Deep learning (DL) techniques for influenza identification are more efficient than traditional approaches in terms of logistics and cost. The benefit of influenza prediction lies in its ability to minimize morbidity and mortality by allowing relevant departments to implement appropriate preventative and control actions after evaluating forecasted data. This research develops a Runge Kutta optimized Dynamic Gated recurrent unit (RKO-DGRU) public health with for influenza identification. Initially, the dataset is collected from kaggle and preprocessed utilizing the lemmatization method. Our approach can result in a sensitivity of 86.69%, specificity of 93.68%, and 97.5% accuracy. The findings highlight the possibility of applying DL approaches to efficiently identify and categorize influenza using data gleaned from conversations on open networks. It can thus provide efficient ways to stop and manage an Influenza epidemic.

1. Introduction

Many vaccination messages on social media express doubts about the safety of vaccines and fabricate patient stories, although the influenza vaccine is one of the greatest successes of public health and avoids millions of diseases and thousands of deaths annually [1]. To minimize the spread of the virus during influenza seasons, early and accurate influenza forecasting is essential. Conventional influenza monitoring often defines and classifies influenza outbreaks that have already occurred since it involves manual data collecting that takes weeks to complete [2]. Regular reports and dissemination of official information on the flu trend are made by government public health organizations, because of the considerable lag in time, these numbers sometimes fall short of providing insight into the most recent evolution of flu outbreaks [6]. Influenza continues to put pressure on the public healthcare system despite persistent attempts to increase vaccination coverage [11]. Influenza is an infectious disease that spreads through the respiratory system and can be fatal, especially in small children and the elderly [8]. It occurs annually in temperate climates as epidemics [4]. There are socioeconomic variations in influenza morbidity and death, according to several researches [3]. Research has demonstrated that influenza hospitalization rates are positively correlated with low levels of education and that the most destitute areas saw twice as many hospitalizations for influenza as the regions with the lowest rates of poverty [5]. A novel RKO-GAN method for influenza identification is developed in this study. The following sections make up the remaining portion of this report: The related work is under portion 2. The method is described in portion 3. In portion 4, the performance evaluations of the article are discussed; in portion 5, conclusions are provided.

Related work

The enhancement of influenza-like illness (ILI) monitoring paradigm, they developed the contactless syndromic surveillance technology known as FluSense Al Hossain et al. [12]. The study examined Duchemin et al. [7] the effectiveness and viability of detecting influenza epidemics utilizing a detection system powered by sick leave data [14]. The paper examined the usefulness of a machine-learned anonymized Venkatramanan et al. [13] mobility map compiled from hundreds of millions cell phones to predict epidemics. They discovered a time-precedence link between influenza epidemics and real-time online data uploads Jang et al. [9].

2. Methodology

The suggested approach RKO-DGRU consists of several parts, such as data collection preprocessing, and classifier. The suggested approach used to identify influenza is shown in Figure 1 every component will be covered in detail in the next subsections.

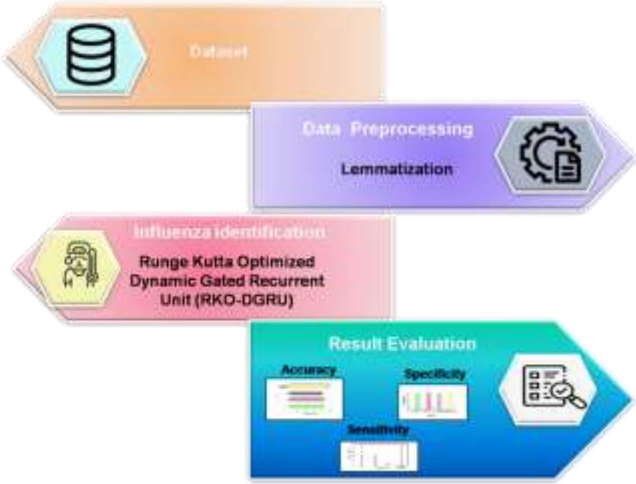

Figure 1. Proposed flow

Dataset

This dataset will examine vaccination as a critical public health intervention in the battle against infectious illnesses. Individuals become immune through vaccinations, and a community's level of immunity can further stop the spread of illness by creating herd immunity. The dataset gathering form (https://www.kaggle.com/datasets/arashnic/flu-data).

Data preprocessing using lemmatization

Several preprocessing processes are carried out to concentrate on the key user post concepts and improve the final feature vector's semantic quality. Remove punctuation, change the text's case to lowercase, and tokenize it before lemmatization. Ensuring precise base word extraction raises the standard of language analysis and comprehension.

Classification using Runge Kutta optimized Dynamic Gated recurrent unit (RKO-DGRU)

Runge-Kutta methods are used in numerical approaches to solve ordinary differential equations. Dynamic Gated Recurrent Units (DGRU) is DL techniques used for regression and classification problems. This indicates the RKO and DGRU were introduced with the suggested approach, which is an advanced and contemporary DL methodology.

Runge Kutta Optimization (RKO)

Three methods are used by the RKO to update the decision variables. It begins by using the RKO technique to determine the location y_{m+1}, which is provided in Equation (1). Next, to raise the caliber of the solutions and prevent local optima stagnation, it employs the enhanced solution method. Equation (2) is used to obtain the new position y_{new2}. The optimal solution is y_{new2} if its fitness is greater than that of y_{m+1}. If not, the expression in Equation (3), which designates a new position called y_{new3}, will be computed. The optimal choice will be y_{m+1} if the cost calculation for y_{new3} is inferior to that of y_{m+1}.

$$y_{m+1} = \begin{cases} (y_v + k \times s \times y_v) \\ +(SF \times SM) + (\mu \times y_g) if rand < 0.5 \\ (y_n + k \times SF \times s \times y_n) \\ +(SF \times SM) + (\mu \times y_g) if rand \geq 0.5 \end{cases} \tag{1}$$

$$y_{new2} = \begin{cases} y_{new1} + k \times u \\ \times |(y_{new1} - y_{avg}) + randn|, if u < 1 \\ (y_{new1} - y_{avg}) + k \times u \\ \times |(w \times (y_{new1} - y_{avg}) + randn|, if u \geq 1 \end{cases} \quad \forall rand < 0.5 \tag{2}$$

$$y_{new3} = (y_{new2} - rand \times y_{new2}) + SF\left(rand \times y_{RK} + (c \times y_p - y_{new2})\right) if rand < u \tag{3}$$

Where μ represents the random number, $randm$ indicates the random number having a normal distribution. k is a randomized value in the region. s is a whole number with a value of $1, 0$ or -1. This option modifies search direction and enhances diversity. The average of the three solutions at random is shown by the symbol y_{avg}. In RKO, SM stands for the primary search mechanism. Because of its randomized adaptive nature, scale factor (SF) helps RKO enhance exploration and exploitation.

Gated Recurrent Unit (GRU)

The basic idea of GRU is to employ gating methods to update a network's hidden state only on a selected fraction of time steps. Gating systems are used to control information entering and exiting the organization. The GRU has two gating mechanisms, the reset gate and the update gate. The special emphasis of this study is time series prediction, for which the recurrent neural network (RNN) approach is a widely used DL technique. But RNNs can also have problems, such as disappearing and gradient explosion, especially as learning long-term relationships from the input. The Long short term memory (LSTM) can alleviate these issues by using a gating mechanism to enhance gradient flow inside a network. A GRU is an LSTM variant that has two gates instead of the LSTM's three gates. The GRU demonstrates an improved ability to identify and learn long-term relationships in time-series data combined with a decrease in computational costs and model complexity, which results in improved training efficiency. This improves the GRU’s suitability for managing time-series data’s long-term dependencies. Due to its reduced storage requirements, the GRU can process huge datasets as well. The major model in this study was decided to be the basic GRU model. Figure 2 shows the architecture of a GRU model.

$$q_s = \sigma(X_q w_s + V_q g_{s-1} + a_q) \tag{4}$$

$$y_s = \sigma(X_y w_s + V_y g_{s-1} + a_y) \tag{5}$$

$$\tilde{g}_s = \tanh\left(X_g w_s + V_g (q_s * g_{s-1}) + a_g\right. \tag{6}$$

$$g_s = y_s * g_{s-1} + (1 - y_s) * \tilde{g}_s \tag{7}$$

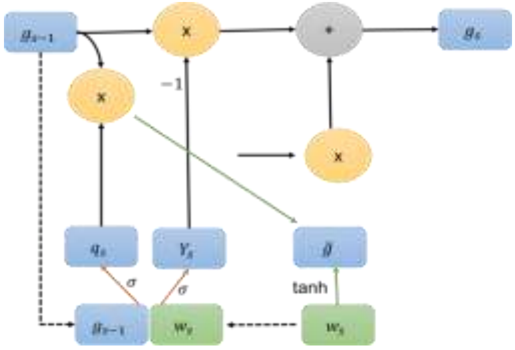

Figure 2. GRU architectural block

As shown in equations (4)-(7), the element-wise product formula is denoted by$*$, The weight matrices of the q_s gate and the y_s gate are denoted by the symbols X_q and X_y, respectively; the weight matrix for the output is represented by V_g. The input data is represented by w_s at time s, the candidate state and output state by $\tilde{g}_s$ and g_s at time s, respectively, by constants a_q, a_y , and a_g, and by the sigmoid and $\tanh$ activation functions, σ and $\tanh$, respectively, that are used to activate the control gates and candidate states.

3. Results and discussion

In this paper, Python 3.11 has been used for identifying influenza trends via social media. A laptop with 32 GB of RAM, an Intel (R) processor, and Windows 10 installed. The quality of the proposed RKO-GAN technique is thoroughly investigated through comparison and evaluation of the outcomes. The effectiveness and precision of a proposed method are contrasted with those of existing techniques such as Boosting trees (BT), Random Forest (RF), and Logistic regression (LR) [10]. The estimated accuracy, sensitivity, and specificity are shown in the result for the provided approach. Accuracy indicates that across many fields, accuracy serves as a fundamental metric for correctness and precision. It illustrates how closely a measured value matches the real value in terms of data analysis. Figure 3 shows the accuracy of the proposed system. The recommended method achieves 95.36%, compared to 84.54% for BT, RF has gained 85.21%, and 81.12% for LR. This confirms that the RKO-DGRU method we proposed is highly accurate.

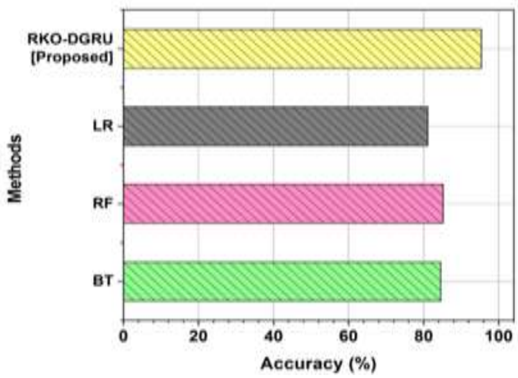

Figure 3. Comparison of the accuracy

The sensitivity of the test is the proportion of disease-bearers that it correctly detects. Sensitivity is the proportion of public healthy persons correctly excluded from the sample by the test. Figure 4 shows the comparative examination of sensitivity using existing methods. The sensitivity for BT has gained 75.64%, RF is offered at 75.59%, LR is at 66.97%, and the highest sensitivity is 86.69% for the suggested RKO-DGRU technique.

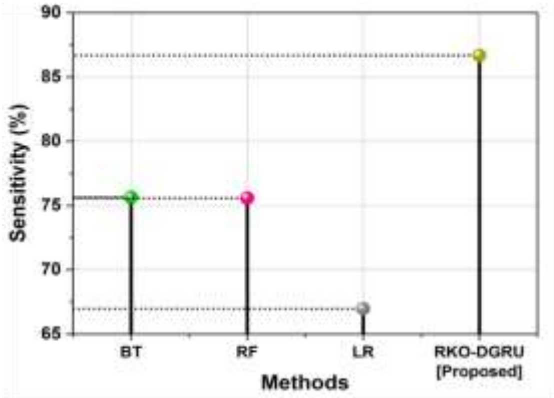

Figure 4. Comparison of the sensitivity

Specificity is the percentage of public healthy individuals that the test successfully eliminates from the sample. These concepts are essential from a clinical perspective to confirm or exclude sickness during screening. Figure 5 shows the accuracy rates for the existing and proposed techniques. The recommended method achieves 93.68% when compared to existing methods, which obtain values of 89.47% for BT, 90.91% for RF, and 89.50% for LR. Table 1 shows the results of the classification techniques in numbers.

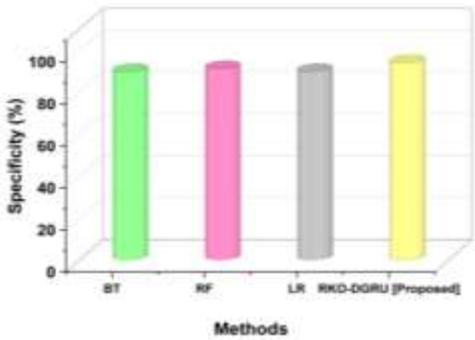

Figure 5. Comparison of the specificity

Table 1. Numerical outcomes of the classification methods

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)
BT	84.54	75.64	89.47
RF	85.21	75.59	90.91
LR	81.12	66.97	89.50
RKO-DGRU [Proposed]	**95.36**	**86.69**	**93.68**

4. Conclusion and future scope

This study proposed a Runge Kutta optimized Dynamic Gated recurrent unit (RKO-DGRU) for identifying influenza trends via social media. Through the GRU with the well-known Runge-Kutta method for differential equation-solving accuracy, this strategy improved the model's capacity to produce realistic synthetic data that mirrors conversations about influenza on social media. The RKO-DGRU enhanced trend identification and prediction accuracy by effectively capturing intricate patterns in temporal data. The results showed a 95.36% accuracy rate, sensitivity (86.69%), and specificity (93.68%) in predicting influenza from the tweets. As for minor influenza complications, sinus and ear infections are widespread of pneumonia, which is a dangerous side effect that can arise from a bacterial co-infection or an isolated influenza virus infection. The goal of future research is to improve the RKO-DGRU model by integrating more data sources and investigating real-time applications for improved public health intervention and influenza trend monitoring.

Reference

[1] M.P.S. Chan, K.H. Jamieson, and D. Albarracin, "Prospective associations of regional social media messages with attitudes and actual vaccination: A big data and survey study of the influenza vaccine in the United States". Vaccine, 38(40), pp.6236-6247, 2020. https://doi.org/10.1016/j.vaccine.2020.07.054

[2] Y. Wang, K. Xu, Y. Kang, H. Wang, F. Wang, and A. Avram, "Regional influenza prediction with sampling Twitter data and PDE model", International journal of environmental research and public health, 17(3), p.678, 2020. https://doi.org/10.3390/ijerph17030678

[3] S. Neelima, Manoj Govindaraj, Dr.K. Subramani, Ahmed ALkhayyat, & Dr. Chippy Mohan. (2024). Factors Influencing Data Utilization and Performance of Health Management Information Systems: A Case Study. *Indian Journal of Information Sources and Services*, *14*(2), 146–152. https://doi.org/10.51983/ijiss-2024.14.2.21

[4] C.M. Zipfel, V. Colizza, and S. Bansal, "Health inequities in influenza transmission and surveillance",PLoS Computational Biology, 17(3), p.e1008642, 2021. https://doi.org/10.1371/journal.pcbi.1008642

[5] S. Amin, M.I. Uddin, M.A. Zeb, A.A. Alarood, M. Mahmoud, and M.H. Alkinani, "Detecting dengue/flu infections based on tweets using LSTM and word embedding", IEEE access, 8, pp.189054-189068, 2020. https://doi.org/10.1109/ACCESS.2020.3031174

[6] Akila, R., & Revathi, S. (2023). Fine Grained Analysis of Intention for Social Media Reviews Using Distance Measure and Deep Learning Technique. *Journal of Internet Services and Information Security*, *13*(2), 48-64.

[7] T. Duchemin, J. Bastard, P.A. Ante-Testard, R. Assab, O.S. Daouda, A. Duval, J.P. Garsi, R. Lounissi, N. Nekkab, H. Neynaud, and D.R. Smith, "Monitoring sick leave data for early detection of influenza outbreaks", BMC Infectious Diseases, 21, pp.1-8, 2021.https://doi.org/10.1186/s12879-020-05754-5

[8] Lomotey, R.K., & Deters, R. (2013). Facilitating Multi-Device Usage in mHealth. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *4*(2), 77-96.

[9] 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), pp.2400-2412, 2021.https://doi.org/10.1109/TNNLS.2021.3106637

[10] R. Kumar, S. Maheshwari, A. Sharma, S. Linda, S.Kumar, and I. Chatterjee, "Ensemble learning-based early detection of influenza disease", Multimedia Tools and Applications, 83(2), pp.5723-5743, 2024.https://doi.org/10.1007/s11042-023-15848-2.

[11] W.N. Harrington, C.M. Kackos, and R.J. Webby, "The evolution and future of influenza pandemic preparedness" Experimental & molecular medicine, 53(5), pp.737-749, 2021, https://doi.org/10.1038/s12276-021-00603-0

[12] F. Al Hossain, A.A. Lover, G.A. Corey, N.G. Reich, and T. Rahman, "FluSense: a contactless syndromic surveillance platform for influenza-like illness in hospital waiting for areas", Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, 4(1), pp.1-28, 2020.https://doi.org/10.1145/3381014

[13] S. Venkatramanan, A. Sadilek, A. Fadikar, C.L. Barrett, M. Biggerstaff, J. Chen, X. Dotiwalla, P. Eastham, B. Gipson, D. Higdon, and O. Kucuktunc, "Forecasting influenza activity using machine-learned mobility map", Nature communications, 12(1), p.726, 2021.https://doi.org/10.1038/s41467-021-21018-5

[14] Bobir, A.O., Askariy, M., Otabek, Y.Y., Nodir, R.K., Rakhima, A., Zukhra, Z.Y., Sherzod, A.A. (2024). Utilizing Deep Learning and the Internet of Things to Monitor the Health of Aquatic Ecosystems to Conserve Biodiversity. *Natural and Engineering Sciences*, *9*(1), 72-83.