Continual Learning for Neural Epidemic Forecasting Models: Final Report

Shikhar Verma
Georgia Institute of Technology
Atlanta, Georgia, USA
sverma87@gatech.edu

Lucy Johns
Georgia Institute of Technology
Atlanta, Georgia, USA
ljohns9@gatech.edu

Siddarth Gulechha
Georgia Institute of Technology
Atlanta, Georgia, USA
sgulechha3@gatech.edu

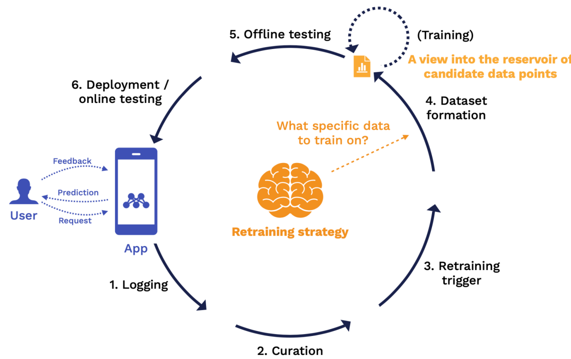

Figure 1: A basic retraining strategy model proposed by "The Full Stack"

Abstract

Accurate and timely epidemic forecasting, particularly seasonal diseases like influenza/flu, is pivotal for public health planning. Traditional methods often require retraining with new data, which can be inefficient and computationally expensive. In order to address this, we propose developing a tool that would integrate continual learning methods to improve flu forecasting. By applying methods such as LSTM, GRU, and TCN, along with techniques like Elastic Weight Consolidation (EWC), Synaptic Intelligence (SI), and Learning without forgetting (LwF), we aim to ensure that the models can update efficiently with new weekly data while retaining knowledge from the past seasons. Using publicly available influenza data from New York, we will test and demonstrate the tool's ability to provide more accurate and reliable predictions, ultimately helping public health officials better manage flu outbreaks.

ACM Reference Format:
Shikhar Verma, Lucy Johns, and Siddarth Gulechha. 2025. Continual Learning for Neural Epidemic Forecasting Models: Final Report. In . ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction and Motivation

In early March 2020, the World Health Organization officially declared COVID-19 a global pandemic. Characterized by a wide range of severe respiratory symptoms, the SARS-CoV-2 virus led to over 10 million deaths worldwide between 2020 and 2021. The pandemic profoundly impacted public health, economies, and daily life, prompting widespread efforts to mitigate its spread and protect vulnerable populations. One significant portion of this widespread effort took the form of innovative epidemiological research, which allowed policymakers and hospitals to formulate prudent decisions and solutions that significantly benefited the general population.

One pivotal tool that was utilized in the research was epidemic forecasting models. These models were and are instrumental in providing actionable insights and predictions that are invaluable in preventing, controlling, and managing disease outbreaks. When more precise, these models allow public health organizations to better prepare for anticipated outbreaks. For instance, they can use the lead time to increase hospital capacity or stock medical supplies, potentially saving lives and reducing the outbreak's impact. Further, policymakers can enforce social distancing measures or mask mandates before the virus cases potentially escalate.

2 Problem Definition

Epidemic forecasting requires machine learning models capable of adapting to evolving disease dynamics while retaining historical knowledge. Unlike traditional machine learning, which trains models on fixed datasets, epidemic forecasting necessitates continual updates as new data becomes available. This poses the dual challenge of learning from new data while avoiding the loss of previously acquired knowledge—a phenomenon known as *catastrophic forgetting*.

Our project focuses on influenza case forecasting using publicly available data from New York, covering weekly influenza case counts during and after the 2009 H1N1 pandemic. Each data point

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

represents a time-series feature set $x_t \in \mathbb{R}^d$, where d includes features like lagged case counts, epiweek indicators, and temporal trends, while the target variable $y_t \in \mathbb{R}$ corresponds to weekly influenza case counts. The dataset spans multiple seasons, allowing the model to learn both seasonal patterns and abrupt changes in disease dynamics.

The formal goal is to predict influenza case counts y_t at time t based on historical data $\{x_1, x_2, \ldots, x_t\}$ by learning a mapping $f_\theta : \mathbb{R}^d \to \mathbb{R}$ that minimizes the prediction error:

$$L(\theta) = \frac{1}{n} \sum_{t=1}^{n} \ell(f_\theta(x_t), y_t),$$

where ℓ is the loss function (e.g., Mean Squared Error). Over time, as new weekly data streams in, the model must be updated to minimize:

$$L_{\text{new}}(\theta) + \lambda R(\theta, \theta_{\text{prev}}),$$

where $R(\theta, \theta_{\text{prev}})$ is a regularization term ensuring the retention of historical knowledge, and λ balances the trade-off between adapting to new trends and preserving prior knowledge.

To achieve these objectives, we propose integrating deep learning models and continual learning methods. Deep learning models are well-suited for capturing temporal dependencies and seasonal patterns from the influenza dataset. By leveraging their capacity to learn complex, non-linear relationships in sequential data, these models ensure accurate predictions of weekly case counts. On the other hand, continual learning methods address the problem of catastrophic forgetting by enabling the model to update itself efficiently with new data without losing historical knowledge. This combination allows the framework to remain computationally efficient while adapting to the dynamic nature of real-world epidemic data. By integrating these two approaches, our project seeks to deliver a robust and adaptive solution for epidemic time-series forecasting.

3 Response to Milestone Comment

Traditional machine learning models typically rely on a fixed dataset divided into two portions, such as a standard 70-30 split, where 70% of the data is used for training, and the remaining 30% is reserved for testing. In contrast, continual learning models operate differently by processing a stream of datasets, enabling the model to "continuously" learn and adapt to new information. In this approach, each dataset in the stream must be split into training and testing portions individually to ensure model accuracy. After training the model on a new dataset, the model is immediately tested to assess its performance. This iterative process, training, and testing after each dataset differs from traditional machine learning models because traditional models train on multiple datasets and test only at the end of the training. However, in continuous learning models, a split-by-patterns strategy evaluates a model's performance for each learning experience rather than a set of experiences. This strategy enables a more accurate measurement of the model's ability to incorporate and apply new information over time.

4 Related Works and Survey

We wanted to get a base idea of the models and methodologies applied in continual learning models. In a study with the Wang lab [7], continual learning is described as a possible machine learning model that is characterized by dynamic distributions. These dynamic distributions constantly change, causing the model to learn for a "lifetime." Since these models are constantly learning new datasets, they may need help to accurately assess older datasets due to adjusting the model to a newer one. Therefore, it is imperative to ensure that continual learning models are robust and have strong generalizability (the model can accommodate various distributions).

Further, Haoxin's study [2] explored the variation that exists within time-series data, data observed at regular intervals. In epidemiology, this type of data is subject to distributional shifts due to underlying factors such as policy intervention or climate change. Because of these possible shifts (also known as out-of-distribution scenarios), it is important to identify parts of the data that will remain consistent so that the model can remain as robust as possible. An invariant learning approach to a model will try to look at the features of the data that remain consistent throughout time intervals, allowing the model to function accurately beyond just the training data.

In Robert Moss' study on seasonal influenza epidemic forecasts [4], their team used Bayesian methods to produce weekly seasonal influenza forecasts. Their forecast method used an SEIR (susceptible-exposed infectious-recovered) population model. The weekly influenza case notification counts produce a median 50% credible interval and 90% credible interval for the future predicted weekly notification counts and timing of the epidemic peak using a particle filter, respectively. Since this model focuses on forecasting seasonal disease, the strength of this model is the introduction of "seasonal forcing," which considers the influence of climate on the spread of influenza. With such modifications, the model can reflect previous influenza seasons' characteristics (timing, duration, and magnitude) before an epidemic signal is present. This strength can improve the accuracy of the forecasting model; however, with the additional climate variables and more complex model, this modification increases computational cost. If we seek efficient continual learning methods, we may need to consider this computational cost with the reduction in forecast error specifically for our data.

Another article provided by Roger A. Morbey [3] introduced their model of study on forecasting respiratory and gastrointestinal infectious diseases with a seasonal component. In this study, they discovered that many previous influenza forecasting models are usually assessed on the accuracy of daily or weekly forecasts, not that of peak activity; therefore, they may not be optimized for forecasting the timing and intensity of seasonal peaks. Based on this observation, they designed a model that performs well when unprecedented seasonal activity improves performance during typical seasons but worsens it for the atypical seasons. Considering "black swan" events in the method means the model can perform effectively even in unusual circumstances. Therefore, the model could enhance the efficiency of forecasting models, reducing computational costs.

Venna et al. [6] in their work proposed an LSTM-based multi-stage forecasting method for influenza, which introduced two novel aspects: capturing the temporal dynamics of seasonal flu and incorporating external variables such as geographic proximity and climatic factors like humidity and temperature. The strength of this work lies in its comprehensive approach to improving influenza forecasting by leveraging both temporal and environmental factors, resulting in better performance than traditional methods. This multi-factor approach offers a promising direction for more accurate epidemic forecasting. However, including external variables increases the complexity of the model, which may hinder its applicability in regions where such environmental data is unavailable or inconsistent. Additionally, while the model performs well, the reliance on external data limits its generalizability across diverse geographical contexts.

Shaghaghi et al. [5] introduce eVision, an LSTM-based tool that forecasts influenza outbreaks using a combination of historical flu data and Google Trends search queries. The key strength of this approach is its practical application in real-time flu forecasting, achieving high accuracy (90.15%) in predicting the 2018–19 flu season in the United States. Using Google search data to capture behavioral trends adds a novel dimension to influenza forecasting. However, the reliance on search data introduces potential noise and bias, as the volume of searches may not always accurately reflect real-world flu cases. Additionally, the study highlights the need for improvements, such as adding confidence intervals and providing state-level predictions to increase the model's reliability and utility in healthcare decision-making.

5 Proposed Method

5.1 Data

Due to the scale of most diseases, we need to obtain a large amount of reliable data from the public healthcare systems that contain the number of incidences, disease prevalence, and cases resulting in death for the model training. For this project, we intend to explore epidemic time-series forecasting models on flu case data in New York. *Influenza* is a contagious respiratory illness caused by influenza viruses and is considered a seasonal disease. New York is a populous city with extensive social activities, which makes the forecasting more significant.

5.1.1 Data Collection. We obtained laboratory-confirmed cases by county of influenza in New York (https://health.data.ny.gov/Health/Hospital-Inpatient-Cost-Transparency-Beginning-200/7dtz-qxmr/about_data), sourced from the official site of the New York State Department of Public Health (NYSDOH). The data set includes clinical laboratory confirmed positive cases of influenza that meet the NYSDOH definition and types of influenza (Influenza A, Influenza B, and Influenza unspecified) by week from 2009 to September 2024. The influenza A label refers to confirmed cases of type A influenza that frequently causes seasonal epidemics, including subtypes such as H1N1 and H3N2. The influenza B label refers to confirmed cases of type B influenza that also contributes to seasonal flu outbreaks but in milder cases. The unspecified influenza label refers to cases in which the specific influenza type (A or B) could not be determined or was not tested during diagnosis. Including the "unspecified" category in analyses helps maintain data integrity, as they might still contribute to the epidemic despite the type of influenza. We considered about joining other data sets that include the death number from influenza, however, the death count is relatively small when counted weekly, thus not very useful in our model study, thus not included at the end.

5.1.2 Preprocessing, Summary Statistics and Initial Findings. Preprocessing is essential for dealing with noisy, incomplete, or inconsistent time series data, especially in public health records. Any mismatches in week numbers or missing value should be filtered and removed. Missing data might occur if certain counties lack reporting in specific weeks.

Based on the general observation of the datasets, one of the initial findings is that the reported infected number is lowest around week 21-39, these weeks are typically considered as the end of the flu season[1]. The low report count suggests that influenza cases naturally decline as the season transitions, we recognize this trend of seasonal decline to the model's seasonality component and removed these weeks with low counts and rearrange the data into Epiweeks 1-33 for the efficacy of our model training.

5.2 Intuition

5.3 Future Ready Considerations

While our current dataset is fixed, we are developing this tool with flexibility for future use cases where data may be updated weekly or seasonally. The continual learning methods and models described here are adaptable to real-time data streams, making this tool suitable for future applications that require ongoing forecasting in response to changing epidemic conditions.

5.4 Algorithms and Models

In this project, we are developing a tool to integrate continual learning techniques into epidemic time-series forecasting, specifically targeting New York influenza cases and death tolls. Our tool will employ a combination of time-series models, continual learning methods, and supporting data processing modules to ensure adaptable, efficient forecasting. Below, we provide a detailed description of each component.

5.4.1 Deep Learning Models for Time-Series Forecasting: The foundation of our tool lies in robust time-series models designed to capture temporal dependencies and seasonal patterns in influenza data. We focus on three core architectures:

a. **LSTM (Long Short-Term Memory Networks)**: LSTM is a type of recurrent neural network (RNN) particularly well-suited for sequential data. It overcomes the limitations of traditional RNNs with its ability to manage long-term dependencies. Each LSTM cell includes mechanisms to retain or forget information over time, making it ideal for capturing trends and seasonal fluctuations in time-series data. In our project, LSTMs will help predict weekly influenza cases and death tolls by learning from previous weeks' data, capturing both short-term spikes and long-term trends.
b. **GRU (Gated Recurrent Units)**: GRU, another type of RNN, serves as a more computationally efficient alternative to LSTM. GRUs simplify the internal structure of LSTMs by combining the forget and input gates, allowing for faster and more efficient training. This architecture is beneficial when computational resources are limited or real-time updates are required. By using GRU, we aim to achieve high accuracy in forecasting with reduced computational demands, making it

suitable for real-time application if our tool were to handle continuous updates.

c. **Temporal Convolutional Networks (TCNs)**: TCNs apply convolutional layers instead of recurrent connections to process sequential data, making them highly efficient for long sequences. TCNs also include mechanisms like causal convolutions, ensuring predictions only rely on past data points. This architecture allows faster training and has shown promising results in time-series applications, often outperforming RNN-based models in certain tasks. By testing TCNs alongside LSTM and GRU, we aim to identify the most effective model for influenza forecasting within our framework.

5.4.2 Continual Learning Methods: Continual learning methods are essential for adapting models to new data while retaining previously learned knowledge. This is particularly important in epidemic forecasting, where new weekly data may reveal emerging trends or changes in disease spread. We explore three continual learning techniques to mitigate the risk of catastrophic forgetting:

a. **Elastic Weight Consolidation (EWC)**: EWC helps the model retain previously learned information by adding a penalty to changes in important parameters during training. This method calculates a "Fisher Information Matrix" to identify crucial weights and selectively protects them from significant updates. In our tool, EWC will allow the model to incorporate new weekly data without forgetting patterns established in earlier flu seasons, ensuring that predictions remain accurate over time.

b. **Synaptic Intelligence (SI)**: Similar to EWC, SI dynamically tracks parameter importance throughout training, but it does so by accumulating changes over time. SI computes the "importance" of each parameter based on how much it contributes to reducing the loss, allowing the model to adapt to new data while preserving essential information. By integrating SI, we aim to give the model flexibility in updating parameters for current trends while ensuring it maintains knowledge of historical trends crucial for long-term predictions.

c. **Learning without Forgetting (LwF)**: LwF leverages knowledge distillation, a technique that uses the predictions of a previous model as "soft targets" for new data, allowing the model to retain older knowledge while adapting to new information. This approach will help us integrate new weekly influenza data into the model without diminishing its performance in earlier periods. By incorporating LwF, the tool will effectively balance historical flu data with recent patterns, avoiding an over-focus on recent data at the cost of forgetting earlier trends.

5.5 Data and Model Management Techniques

To streamline data handling and optimize model training, our tool will include the following preprocessing and management techniques:

a. **Data Cleaning and Normalization:** Our tool will standardize data formats, handle missing entries, and normalize features such as weekly cases and death tolls, ensuring consistent input for the models.

b. **Temporal Alignment:** Our tool will align data across different time frames to provide seamless input for time-series models.

c. **Model Saving and Reloading:** Given the continual learning nature of this project, model checkpoints and versioning are essential. We plan to implement efficient model saving and reloading mechanisms, allowing researchers to preserve the model state at different time points or experiment with different continual learning techniques without retraining from scratch.

6 Experiments and Results

We implemented three deep learning models. Considering the dataset, we used data from 2009-2021 as the train set and 2022 as the test parameters. We include metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Pearson Correlation Coefficient (PCC). These metrics will enable researchers to evaluate forecasting accuracy.

Table 1: Performance Comparison of Deep Learning and Continual Learning Methods

Deep Learning	**LSTM**	**GRU**	**TCN**
MAE	292.32	4322.98	1760.08
RMSE	508.54	7209.63	3645.52
PCC	0.9998	0.8753	0.9927
Continual Learning	**EWC**	**SI**	**LwF**
MAE	567.52	5497.84	732.72
RMSE	1075.48	7823.22	1398.10
PCC	0.9997	0.9220	0.9988

6.1 Measuring Accuracy

A machine learning model's accuracy is often determined by its ability to align its predictions with the actual outcomes that occur within the population. We present the following techniques to measure the model's ability to accurately predict H1N1 virus cases:

a. **Time Split Validation Method**: This method will split the data used on the model into earlier time points and later time points. We trained the model on the earlier time points (from 2009 to 2021) as historical train set, and test the model's accuracy on the later time points (2022). This will allow us to measure the model's ability to make real-world predictions and how it can stay consistent with changing data.

b. **Evaluation Metrics**: Evaluation metrics can be used to assess how well the model fits to the ground truth. We can use different metrics to assess how well the model can fit, including:
   - *Root Mean Squared Error (RMSE)* - mean squared error will evaluate the average differences between the values predicted by the model and the ground truth values (the square root will just make those values more readable). Assessing this value at specific time steps will help us determine how well the model can predict over time and how robust it is when we provide varying datasets.

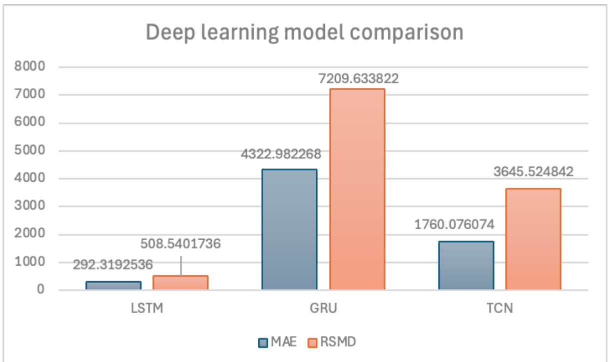

Figure 2: MAE and RMSD Comparison for Deep Learning Models

- *Mean Absolute Error (MAE)* - mean absolute error measures the average magnitude of errors between predicted and actual values, providing an intuitive, direct measure of forecasting accuracy. Calculating MAE at specific time steps will help us understand how consistently the model performs over time, and allow us to compare MAE and RMSE of model performance difference between new tasks and past tasks in order to evaluate the performance of continual learning methods.
- *Pearson Correlation Coefficient (PCC)* - this coefficient value is a number between -1 and 1 that can model the correlation between two data sets. We can assess the PCC of the ground truth and the model's predictions and see how well they correlate, giving us an idea of its accuracy. Here's the formula to calculate that value:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where - x and y are the two variables, - x_i and y_i are the individual values, - $\bar{x}$ and $\bar{y}$ are the means of x and y, respectively.

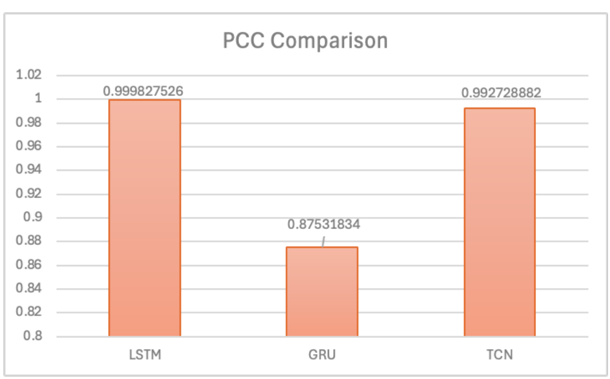

Figure 3: PCC Comparison for Deep Learning Models

6.2 Experiments and Analysis

(1) **LSTM Ensemble with RandomForestRegressor**

(a) **Model and Approach** - The LSTM ensemble model with RandomForestRegressor effectively captured the complex temporal dynamics of weekly influenza case counts. By leveraging the LSTM's ability to model sequential patterns and combining it with RandomForestRegressor to correct residual errors, the ensemble demonstrated its capacity to align closely with observed trends. This approach enabled the model to accurately reflect the sharp peak in cases during the flu season and the gradual decline thereafter.

(b) **Inferences and Observations** - The ensemble model's robust performance highlights its effectiveness in forecasting tasks involving temporal data. The combination of LSTM and RandomForest leveraged complementary strengths, resulting in reliable predictions. However, minor deviations during specific weeks suggest potential areas for further refinement, such as exploring additional features or optimizing hyperparameters. This method underscores the importance of hybrid approaches in modeling complex public health datasets.

(2) **LSTM with EWC**

(a) **Model and Approach** - The Elastic Weight Consolidation (EWC) model with LSTM as the base architecture was designed to address the challenges of continual learning for influenza case prediction. EWC effectively mitigates catastrophic forgetting by penalizing significant updates to weights important for previously learned tasks. By incorporating influenza case counts and epiweeks as features, the model leveraged temporal dependencies to adapt predictions to evolving patterns.

(b) **Inferences and Observations** - The EWC model exhibited strong performance in capturing overall trends and seasonal variability in the data. Despite variability in predictions across different runs due to stochasticity in weight initialization and EWC penalties, the model consistently aligned well with observed patterns, highlighting its robustness in continual learning scenarios. This method emphasizes the potential of EWC in maintaining performance on earlier learned data while adapting to new information, making it suitable for dynamic, time-dependent datasets like influenza forecasting.

(3) **SI with GRU Hidden Layer**

(a) **Model and Approach** - Generally, GRUs tend to perform better on simpler, shorter datasets, while LSTMs excel in handling more complex, variable data. Hence, we sought out to observe this difference by utilizing this deep learning model as a base in Synaptic Intelligence. During training, synaptic intelligence initially monitors the importance of parameter by evaluating its contribution to the loss over time. Once a new task is given, it optimizes the weights based on the previous task given and adjusts parameters according to each new loss function.

Inferences and Observations - Given that the H1N1 dataset is highly variable and complex, this likely explains why the SI model yielded suboptimal results compared to the other models. This is likely due to the hidden layer being used in the continual learning model: GRU. The GRU model can struggle with complex, dynamic data because the model excels more in stable, uniform data. As the training dataset was expanded by adding more years, both the Mean Absolute Error (MAE) and Root Mean Squared Deviation (RMSD) worsened. This suggests that as the dataset grew, the model encountered more variability, particularly in the late 2010s, which the simpler GRU architecture struggled to adapt to. Consequently, the continual learning model had difficulty handling the evolving data, leading to declining performance over time. In comparison to the base model, it seemed the continual learning model worsened in comparison to the deep learning model, implying that utilizing continual learning models can potentially worsen predictions.

(4) **TCN and LwF**
  (a) **Model and Approach** - The TCN model implemented with LwF leveraged the strengths of continual learning for influenza forecasting. By freezing the base TCN layers trained on historical data (2009-2020) and introducing a new output layer for fine-tuning on recent data (2021), the model preserved past knowledge while adapting to new patterns. This approach enabled efficient updates with minimal retraining.
  (b) **Inferences and Observations** - The implementation improved the model performance, as shown, MAE was reduced by 54% and RMSE was reduced by 67% compared to the traditional model, indicating the continual learning method was able to predict the actual count better. Both models have high PCC indicating both models capture the overall trends and patterns accurately, and with a slight increase in PCC with the LwF-enhanced model suggesting improved trend adherence as well. Results shows the benefits of continual learning using LwF, by leveraging the knowledge from the previously trained TCN model (2009-2020), LwF can fine-tune on new data (addition of 2021 dataset). This result highlights LwF's potential for enhancing performance in dynamic, temporal forecasting tasks for public health data.

6.3 Performance Monitoring and Evaluation

To provide researchers with actionable insights, our tool includes built-in performance monitoring for each continual learning model:

a. **Evaluation Metrics:** We include metrics such as Mean Absolute Error (MAE), Root Mean Squared Deviation (RMSD), and Pearson Correlation Coefficient (PCC). These metrics enable researchers to evaluate forecasting accuracy, monitor model efficiency, and observe how well the model adapts to new data without forgetting past trends. Additionally, these metrics allow a detailed comparison of different continual learning methods, such as Elastic Weight Consolidation (EWC), Synaptic Intelligence (SI), and Learning without Forgetting (LwF), in tracking both short- and long-term prediction accuracy.

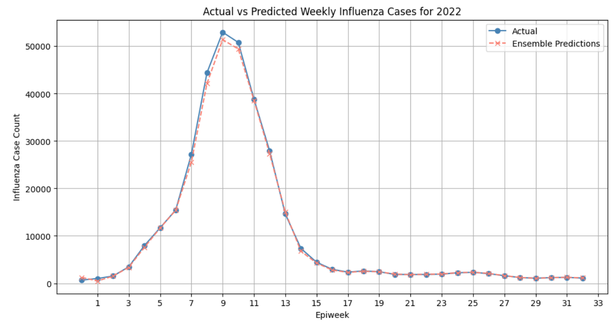

Figure 4: LSTM Ensemble Method

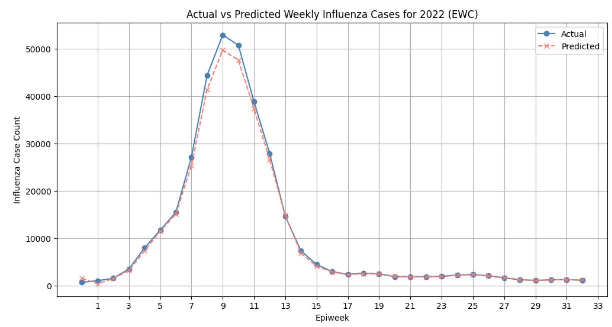

Figure 5: EWC

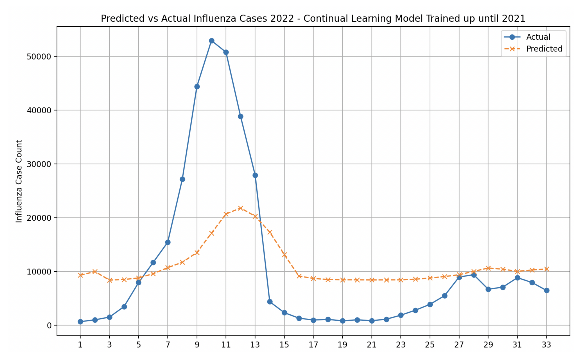

Figure 6: SI

b. **Visualization:** Graphical tools provide a clear visual comparison of model predictions against actual case data, enabling a deeper understanding of model behavior. Visualizations are particularly critical for evaluating continual learning techniques. Below are the prediction graphs for LSTM-based Ensemble Method, EWC, and LwF, highlighting the alignment between actual and predicted influenza case counts.

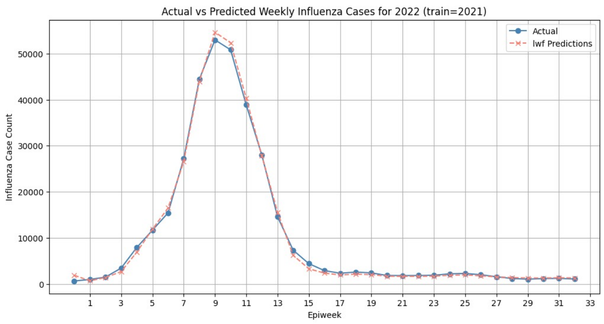

Figure 7: LwF

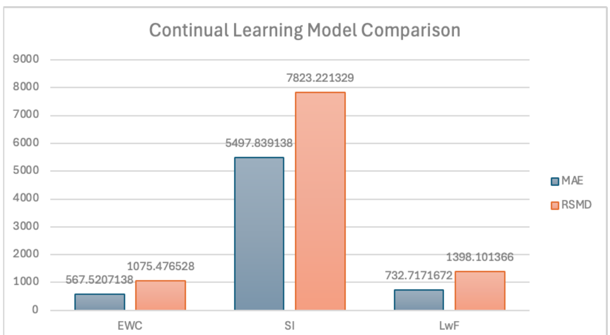

Figure 8: MAE and RMSD Comparison

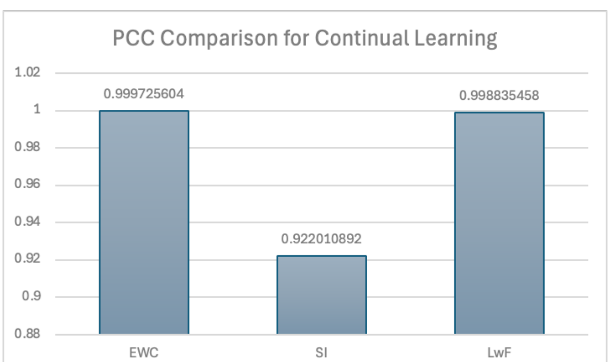

Figure 9: Comparison of Continual Learning Models in terms of PCC

Table 2: Performance Metrics for Continual Learning Models

Model	MAE	RMSD	PCC
LSTM Ensemble	336.63	646.38	0.9997
EWC	434.36	820.00	0.9995
LwF	398.45	702.10	0.9988

7 Conclusion and Further Investigation

7.1 Conclusion

In this study, we conducted a comprehensive evaluation of LSTM, GRU, and TCN models for influenza forecasting. The results demonstrated that the LSTM model outperformed the other architectures, achieving the highest predictive accuracy. This superior performance is likely attributed to LSTM's ability to effectively capture complex temporal dependencies and patterns in sequential data. By contrast, GRU exhibited lower predictive accuracy, reflecting its simpler architecture, which can struggle with datasets requiring long-term memory retention. TCNs, while competitive, showed slightly lower accuracy than LSTMs, possibly due to their reliance on convolutional operations that may not fully exploit sequential dependencies.

Among the continual learning approaches examined, **LwF** emerged as the most effective method. LwF demonstrated a remarkable ability to retain knowledge from previously seen data while simultaneously adapting to new datasets. By leveraging prior task information, LwF mitigated the risk of catastrophic forgetting, allowing the models to maintain robust performance even when trained incrementally on updated influenza case data. This adaptability is particularly important in dynamic, time-dependent scenarios like influenza forecasting, where new data is frequently incorporated. Conversely, **SI** and **EWC** provided only limited improvements. Both methods focus on preserving important model parameters from earlier tasks by penalizing significant weight changes, which helps prevent catastrophic forgetting. However, the effectiveness of these approaches was constrained in this study, likely due to their limited ability to accommodate the temporal complexity and variability of influenza data.

7.2 Further Investigation

Further investigations revealed that updating the base model annually with new data yields more accurate predictions compared to longer update intervals. Our findings showed that shorter update cycles effectively balance knowledge retention and adaptation, as evidenced by the performance metrics across various continual learning approaches. For instance, LwF achieved the highest improvement over time when trained with datasets incrementally extended from 2009–2019 to 2009–2021, as shown in the attached figures. This highlights the importance of leveraging continual learning strategies to dynamically integrate new information while preserving historical trends.

7.3 Discussion and Future Work

Future work can explore integrating external datasets such as weather patterns, vaccination rates, and mobility data to improve model robustness further. Additionally, incorporating advanced continual learning techniques, such as meta-learning or dynamic weight allocation, could potentially address the limitations observed with SI and EWC. Exploring real-time forecasting and adaptive retraining frameworks would also enable faster responses to sudden outbreaks, supporting proactive public health management. Overall, our study highlights the effectiveness of integrating LSTM with

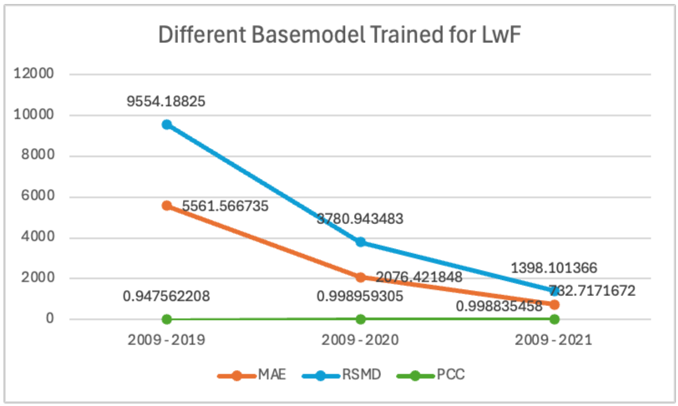

Figure 10: Performance of Different Base Models Trained for SI

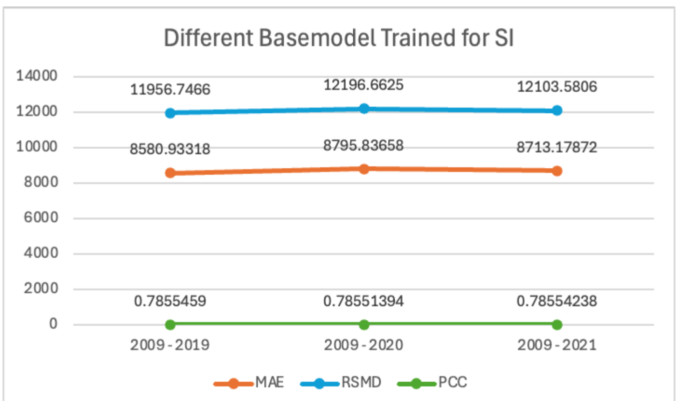

Figure 11: Performance of Different Base Models Trained for LwF

LwF for influenza forecasting, providing a foundation for future advancements in dynamic epidemic prediction models.

Further, we learned that it also may be a beneficial idea to test a continual model's ability to train when the model is fed more initial data. To frame this, we could feed the model the data in increasing increments and monitor how long it takes the model to produce an amount. We could surmise that if the model takes a longer amount of time to train and test the data we're passing, it would indicate that the model performs worse in comparison to other counterparts.

References

[1] Centers for Disease Control and Prevention. 2023. Frequently Asked Flu Questions: 2023-2024 Influenza Season. https://www.cdc.gov/flu/about/season.html Accessed: 2024-11-03.

[2] Haoxin Liu, Harshavardhan Kamarthi, Lingkai Kong, Zhiyuan Zhao, Chao Zhang, and B. Aditya Prakash. 2024. Time-Series Forecasting for Out-of-Distribution Generalization Using Invariant Learning. In *Forty-first International Conference on Machine Learning*. https://openreview.net/forum?id=SMUXPVKUBg

[3] R. A. Morbey, D. Todkill, C. Watson, and A. J. Elliot. 2023. Machine learning forecasts for seasonal epidemic peaks: Lessons learnt from an atypical respiratory syncytial virus season. *PLoS One* 18, 9 (Sep 2023), e0291932. https://doi.org/10.1371/journal.pone.0291932 Published on Sep 22, 2023.

[4] R. Moss, A. E. Zarebski, P. Dawson, L. J. Franklin, F. A. Birrell, and J. M. McCaw. 2019. Anatomy of a seasonal influenza epidemic forecast. *Communicable Diseases Intelligence (2018)* 43 (Mar 2019). https://doi.org/10.33321/cdi.2019.43.7 Published on Mar 15, 2019.

[5] N. Shaghaghi, A. Calle, and Y. Qian. 2020. eVision: Influenza Forecasting Using CDC, WHO, and Google Trends Data. In *2020 IEEE / ITU International Conference on Artificial Intelligence for Good (AI4G)* (Geneva, Switzerland). 38–45. https://doi.org/10.1109/AI4G50087.2020.9311072

[6] S. R. Venna, A. Tavanaei, R. N. Gottumukkala, V. V. Raghavan, A. S. Maida, and S. Nichols. 2019. A Novel Data-Driven Model for Real-Time Influenza Forecasting. *IEEE Access* 7 (2019), 7691–7701. https://doi.org/10.1109/ACCESS.2018.2888585

[7] L. Wang, X. Zhang, H. Su, and J. Zhu. 2024. A Comprehensive Survey of Continual Learning: Theory, Method and Application. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 8 (Aug 2024), 5362–5383. https://doi.org/10.1109/TPAMI.2024.3367329