DSC 180B
UCSD, Winter 2023

Optimizing DeepGLEAM Model for Flu Prediction

Xiangyi Kong, Rhee Kang, Justin Phen
University of California San Diego
{xkong,rkang,jpphen}@ucsd.edu

1 Abstract

The current COVID-19 pandemic and common flu highlight the importance of time-sensitive information in biomedical institutions, politics, and economics. The application of data science in creating real-time predictive models is crucial to help researchers and world leaders better understand disease spread and take preventative measures. Our project aims to optimize the DeepGLEAM model, a deep learning and stochastic process-based predictive model, by better processing raw data and simulating it for improved flu case predictions in the United States. We close the gap in missing data in the DeepGLEAM model by researching interpolation and imputation techniques to enhance the model's predictions. The performance of our adjusted model will be compared against a control model for effectiveness.

2 Introduction

Due to the current state of affairs, the importance of time-sensitive information, especially regarding the livelihood and health of countries across the world has come to the forefront of biomedical institutions, as well as in politics and economics. Data science, in its essence, derives necessary analysis and insights amalgamated from computational programming, as well as a fundamental understanding of statistics and decision-making on a technical level. Consequently, data science is crucial in its application in a vast array of fields and issues, including the current COVID-19 endemic and the common flu, being able to create real-time predictive models that would enable researchers and world leaders to be more aware of the spread and preemptive measures against disease. Therefore, within our project proposal for Winter Quarter 2023, we aim to fully grasp the mechanics of Professor Yian-Ma and Professor Yu's DeepGLEAM model and further optimize its ability to merge simulated and ground truth data so it better predicts future cases of flu cases within the United States.

Deep learning and stochastic processes can be very effective in better understanding and forecasting future events relevant to our society, one of which is the recent endemic COVID-19, as well as the common flu. The concept of predictive distributions through previous truths and predicting with a relative uncertainty projected into the future is quite compelling, for it applies spatiotemporal insights through quantifiable evidence within code, such as residuals. Because disease is constantly changing, mutating, and affecting across state fronts, weekly data predictions are not only necessary but cumulative in providing cognizance in how illness can permeate and any patterns we as data scientists can derive from the perpetual change. Our task mainly lies in how we are able to better process the provided raw data and simulate it so that the forecasting of cases align more closely with future predictions.

The issue with missing data is more inaccurate predictions and its proclivity to increasing uncertainty in predictions. In order to provide our personal contribution towards the already functional DeepGLEAM model, we as a team have researched how to possibly 1) interpolate missing data through a feed-forward system, 2) understand the weight vectors of hyperparameters when splitting the model, and 3) considering other possibilities of imputing the missing data, simply due to most states lacking at least a couple weeks of information from the original shape provided within the flu data file. We aim to take the baseline replication of the DeepGLEAM model from Quarter 1, train individual models per each state's data without tampering with any of it, altering the data through interpolation and imputation, and finally juxtapose the results of the adjusted model's ability to predict future cases against the control model.

3 Methods - Experiment Design

For flu forecasting, we used DSMLP (UC San Diego's Data Science/Machine Learning Platform) for training and testing models. In the case of forecasting without interpolation, the limited availability of GLEAM simulation data resulted in a 0.67:0.33 split between the training, validation, and testing sets. Meanwhile, in the case of forecasting with interpolation, a 50:50 split was utilized between the training, validation, and testing sets. For interpolation, we tried randomly selecting a portion of the real data and fill in the gaps in the dataset. We utilized 1000 epochs and a batch-size of 6 for forecasting without interpolation and 1000 epochs and a batch-size of 11 for forecasting with interpolation. The ADAM optimizer was utilized with a learning rate of $1e^{-2}$ for both cases, and early stopping was implemented to prevent over-training. The input sequential length for the DeepGLEAM Flu model was 1.

For both the Exponential Smoothing (ETS) algorithm and the Auto Regressive Integrated Moving Average (ARIMA) model, we utilized CDC flu ground truth data after 2020-10-10. The parameters for each state were optimized based on grid search results to improve performance.

We also adopted a Multi-Layer Perceptron (MLP) model to tackle the flu forecasting problem, training on normalized data separately for each state. The MLP model was designed with three hidden layers, an input length of 6, an output length of 4, and a hidden layer length of 256. To optimize the model, we employed the Adam optimizer with a learning rate of 0.001, along with a stepwise learning rate decay using the StepLR scheduler. As the loss function, we chose the mean squared loss function. The training process was carried out for a total of 300 epochs, with the training loss and validation loss being reported every 10 epochs. The optimal model was selected based on the lowest validation loss achieved during the training process.

For the combined method, we integrated both the MLP and DeepGLEAM approaches. In the individual MLP, the model was initialized with an input length of 6, an output length of 1, and a hidden layer length of 128. A distinct predictive model was developed for each week, leveraging information from the preceding 6 weeks.

This methodology provides a comprehensive and robust approach to solving the flu forecasting problem, incorporating both MLP and DeepGLEAM techniques to enhance the accuracy and reliability of the predictions. By adopting a systematic training process and refining the model based on validation loss, we ensure the selection of the best-performing model for the task at hand.

4 Methods - Experiment Results

Based on the results depicted in Figure 1, it is evident that the GLEAM prediction dataset presents a significant gap in the records from June 2022 to November 2022. However, after performing interpolation on the dataset, as presented in Figure 2, we were able to impute the trend of flu cases and include an additional five months of data into the available dataset for DeepGLEAM. The improved performance of DeepGLEAM post-interpolation is demonstrated in Table 1, where an average MAE improvement of about 4 is observed. The percentage improvement in MAE when using the Quantile model with GLEAM interpolation compared to the Quantile model without interpolation is approximately 11.85%. Therefore, the utilization of interpolation techniques in this study has not only enhanced the accuracy of the DeepGLEAM model but has also allowed for the inclusion of additional data points, thereby strengthening the reliability of the predictions.

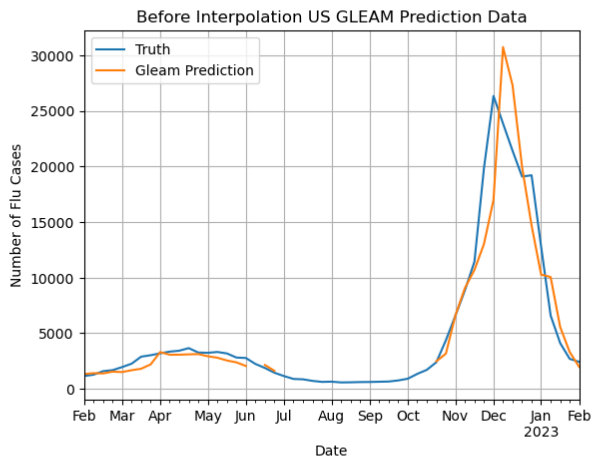

Figure 1: GLEAM Prediction and Truth Before Interpolation

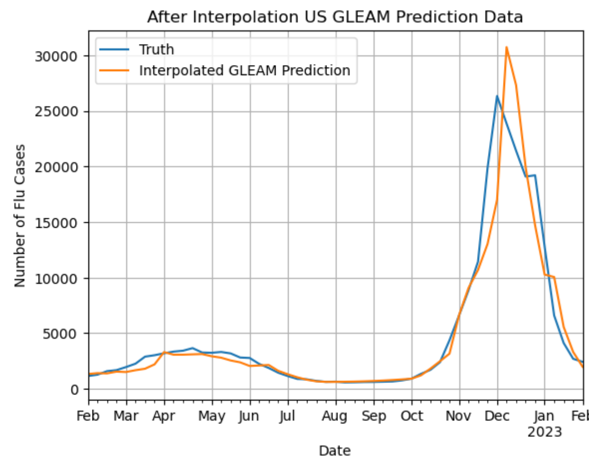

Figure 2: GLEAM Prediction and Truth After Interpolation

Flu Incident					
Model	MAE	Weeks Ahead	Training Time	Epoch Number	seqlen
Quantile	34.64	1	1-2 mins	epo62	1
Quantile *	30.53	1	1-2 mins	epo170	1

*With GLEAM Interpolation

Table 1: Comparison of Autoregressive DeepGLEAM Model performance for Flu forecasting with and without interpolation

Based on the findings presented in Table 2, when the prediction is shifted leftward by 2 weeks (indicated by *), there is an improvement in the model's performance. Specifically, for a one-week-ahead prediction, the MAE is reduced from 20.66 to 17.00, and for a two-week-ahead prediction, the MAE decreases from 23.37 to 22.27. This improvement can be attributed to the delayed nature of the data provided, as shifting the prediction leftwards helps mitigate the effect of data delay. By doing so, the model is better equipped to provide more reliable and accurate predictions, ultimately enhancing its overall performance. The table demonstrates that incorporating a 2-week leftward shift in the Autoregressive Multi-Layer Perceptron model leads to improvements in flu forecasting accuracy, as evidenced by the reduced MAE values.

Flu Incident					
Model	MAE	Weeks Ahead	Training Time	Epoch Number	Learning Rate
MLP	20.66	1	1 min	epo31	0.001
MLP *	17.00	1	1 min	epo31	0.001
MLP	23.37	2	1 min	epo31	0.001
MLP *	22.27	2	1 min	epo31	0.001

MLP: Multi-Layer Perceptron
*Left Shifted by 2 weeks

Table 2: Comparison of Autoregressive Multi-Layer Perceptron Model for Flu forecasting with and without shifting.

As shown in Figure 3 and Figure 4, the performance of time-series forecasting models, such as ETS and ARIMA, are impacted by a lack of trend and limited data points. In the absence of a clear trend, these models struggle to capture the underlying patterns and produce reliable predictions. Additionally, with limited data points, the models do not have enough information to identify and account for seasonality or other factors that can influence the outcome being predicted. As a result, the performance of ETS and ARIMA models suffer under these conditions, highlighting the importance of having a sufficient amount of high-quality data when using these methods for time-series forecasting.

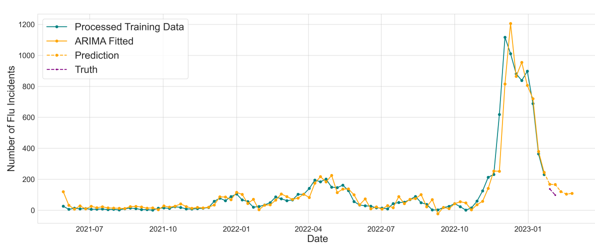

Figure 3: GLEAM Prediction and Truth Before Interpolation

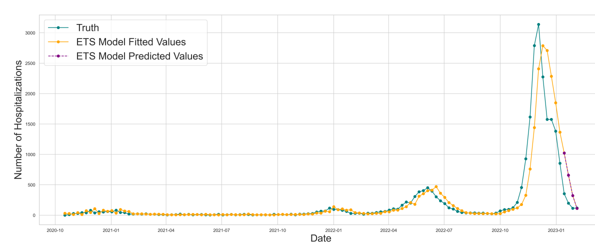

Figure 4: GLEAM Prediction and Truth After Interpolation

Table 3 presents a comparison of the performance of various methods for time series forecasting, and it is evident that the combined MLP and DeepGLEAM approach outperforms all other methods, demonstrating a significant lead in MAE error. This result is a testament to the effectiveness of combining multiple techniques to enhance the accuracy and reliability of time series forecasting models. The success of this approach can be attributed to the complementary strengths of MLP and DeepGLEAM, which when combined, are capable of capturing a wider range of patterns and relationships within the data. Ultimately, the superior performance of the combined MLP and DeepGLEAM approach underscores the importance of leveraging multiple methods and techniques to produce more accurate and reliable predictions in time series forecasting.

Horizon H	MLP	DeepGLEAM	GLEAM	ARIMA	ETS
1W	**10.90**	13.93	12.67	28.6	58.42
2W	17.37	16.05	**15.23**	31.5	60.26
3W	15.71	**14.75**	15.03	33.8	61.53
4W	15.22	**14.09**	15.13	34.9	64.75

Table 3: MAE comparison of different approaches for Influenza Hospitalization forecasting

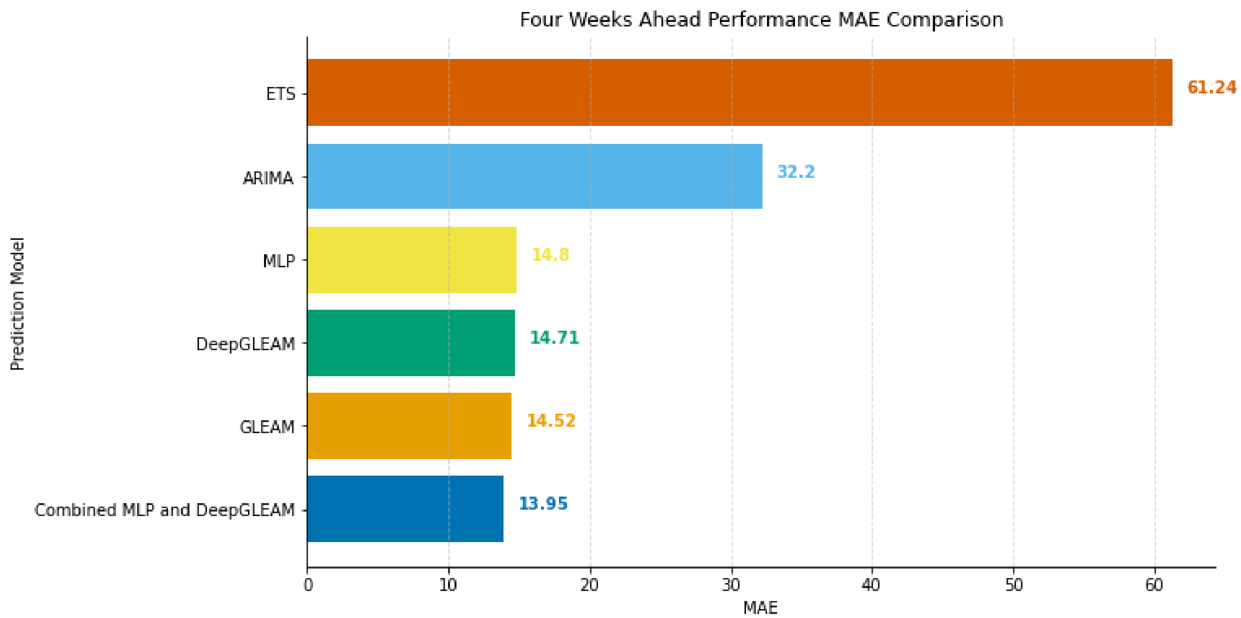

Figure 5: Four weeks ahead performance MAE comparison

Figure 6 depicts the results of a comparison between MLP and DeepGLEAM models for temporospatial forecasting of flu cases in 10 selected states out of the 50 in the United States. The findings indicate that MLP tends to perform better in the short term, particularly within the first and second week of forecasting. In contrast, DeepGLEAM exhibits superior performance in the long term, particularly within the 3-4 week range. The gray area in the graph represents the 95 percent confidence interval and provides an indication of the level of uncertainty associated with the predictions. The results of this study underscore the importance of selecting the appropriate model for the specific time frame being forecasted, as different methods may have varying strengths and limitations depending on the time horizon being considered.

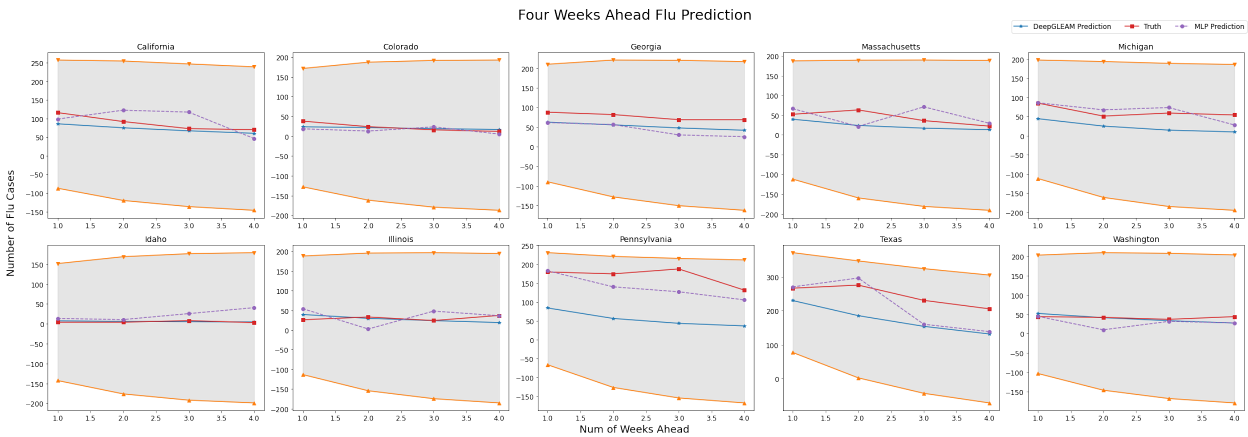

Figure 6: Four Weeks Ahead Flu Prediction

5 Conclusion

Upon transitioning from multiple approaches of interpolation, ranging from simple pandas libraries to more advanced neural networks, we have come to realize that the Multi-Layer Perceptron (MLP) merged with DeepGLEAM provides the best model performance, even outperforming the default model provided in the GLEAM research paper.

Some further steps we would like to consider is the possibility of cross-validating this new merged model back to the COVID-19 dataset, among other endemic and pandemic time-sensitive information. We also would like to consider the Bayesian Information Criterion, as opposed to the Akaike Information Criterion provided in the paper, due to the tendency for most of the models to be Bayesian. Additionally, we can re-evaluate our standard of metric to more than simply MAE, because although Mean Average Error is holistic, it might not be as comprehensive as other metrics. We would also be interested in exploring more complex neural networks, beyond the Multilayer Perceptron, in hopes of further optimizing the DeepGLEAM model. Possibly creating a more thorough dataset, with more data points, could aid in the process of improving model performance with a bigger pool for model training and validating.

We believe that our results and purpose within this capstone project is impactful for the future of deep learning in the healthcare industry, and we hope to find more progress through further exploration of time-series data.