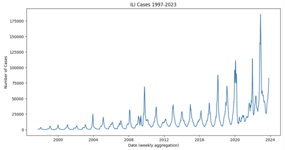
Shifting Patterns: Investigating Recent Changes in Influenza Seasons

Bella Karduck and Theodora Vorias

Background

- There are upwards of 1 billion annual cases of influenza, with as many as 650,000 respiratory deaths each year
- Historically, annual flu seasons have been relatively stable in of number of cases, severity, and duration of the season
- Over the past 5 years, flu seasons have been uncharacteristically severe and early



Research Questions:

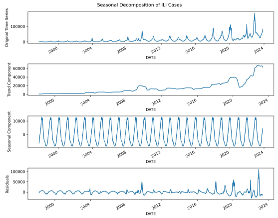
1. Are recent annual flu seasons anomalous as compared to previous seasons?
2. Can models more accurately predict upcoming flu seasons by only training on more recent flu seasons?

Data

- The data is sourced from the Center for Disease Control and Prevention's U.S. Outpatient Influenza-like Illness Surveillance Network (ILINet)
- Weekly data on number of ILI cases across the US from 1997-2023

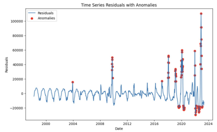
Time Series Analysis

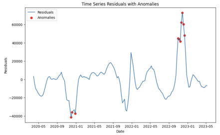
- To investigate seasonal pattern changes in recent years, we conducted a time series analysis, including seasonal decomposition
- Using an Augmented Dickey-Fuller Test, we assessed the stationarity of the data across two periods: 1997-2023 and 2019-2023



Anomaly Detection

- We used the IQR method (k=1.5) to detect anomalies in the residuals of both the full time series (1997-2023) and the COVID-present time series

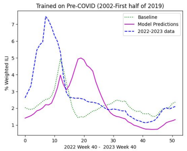


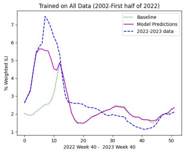


Modeling

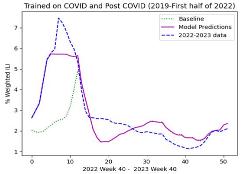
We split the flu season data into 3 time frames:

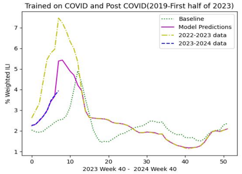
- Pre-COVID-19 **(2002-03 season - 2018-19 season)**
- Pre-COVID-19 through Post-COVID-19 **(2002-03 - 2021-22)**
- COVID-19 and Post-COVID-19 **(2019-20 - 2021-22)**





- Our baseline is the 2021-22 season and to test performance, we predicted on the most recent full flu season, 2022-23
- From a handful of models initially tested, we chose to fine-tune XGBoost models to predict % weighted ILI (Influenza Like Illness) on the 3 time frames and on the upcoming flu season (2023-24)





Results

- For both the full time series and the more recent years time series, the p-value in the ADF test was below 0.05 (p-value > 0.00 for both), so we rejected the null hypothesis and concluded that they are stationary.
- However, a visual analysis of their residuals shows a clear pattern of higher spikes in more recent years.
- The number of outliers greatly decreased between the full time series and the more recent years.
- Training on Pre-COVID-19 data is not effective for predicting current flu patterns; it performed worse than the baseline (RMSE=1.954)
- Models that were trained on data from COVID-19 and on performed best (RMSE=0.597), showing the effect of concept drift

Conclusions

- We should anticipate high variability in upcoming flu seasons in terms of severity, number of cases, and time of peak cases as compared to pre-COVID seasons
- For the most accurate modelling of percent weighted ILI, only the most recent years should be used and models need to be consistently updated as new weekly data is released
- Future directions for U.S. public health officials to consider:
  - Analyzing Post-COVID-19 trends at the state and/or age level
  - Adjusting flu prevention strategies to account for temporal shifts in peak sickness