A Novel Approach for Time Series Forecasting of Influenza-like Illness Using a Regression Chain Method

Nooriyah Poonawala-Lohani[†,1], Patricia Riddle[1], Mehnaz Adnan[2], Jörg Wicker[1]

(1) School of Computer Science, University of Auckland,
Auckland 1010, New Zealand
(2) The Institute of Environmental Science and Research (ESR),
Porirua 5022, New Zealand
[†]E-mail: n.poonawala@auckland.ac.nz

Influenza is a communicable respiratory illness that can cause serious public health hazards. Due to its huge threat to the community, accurate forecasting of Influenza-like-illness (ILI) can diminish the impact of an influenza season by enabling early public health interventions. Machine learning models are increasingly being applied in infectious disease modelling, but are limited in their performance, particularly when using a longer forecasting window. This paper proposes a novel time series forecasting method, Randomized Ensembles of Autoregression chains (REACH). REACH implements an ensemble of random chains for multistep time series forecasting. This new approach is evaluated on ILI case counts in Auckland, New Zealand from the years 2015-2018 and compared to other standard methods. The results demonstrate that the proposed method performed better than baseline methods when applied to this ILI time series forecasting problem.

Keywords: Influenza; time-series; machine learning; forecasting

1. Introduction

A report by the World Health Organization (WHO) shows that each year about 1 billion cases worldwide with up to 650,000 deaths globally are caused by respiratory diseases from seasonal influenza.Influenza poses one of the greatest public health challenges in our world today [a] In New Zealand 10-20 percent of the population are affected by Influenza each year which can be overwhelming for healthcare providers during winter epidemics. Seasonal Influenza generally tends to occur each year during the winter months and spread upon contact. This virus can cause illness that can range from mild to severe and in some cases even cause death. The seasonal influenza virus evolves constantly which is why the spread needs to be monitored, and the vaccines are reviewed twice yearly.

The term ILI will be used throughout the paper to abbreviate Influenza like illness. Weekly ILI surveillance in New Zealand is currently carried out by collecting data from sentinel General Practices (GPs), laboratories, a consumer helpline (Healthline) and emergency departments. Figure 1 is a snapshot of an ILI surveillance dashboard used in New Zealand.

Recent findings by the WHO states that the world is not prepared to respond to an

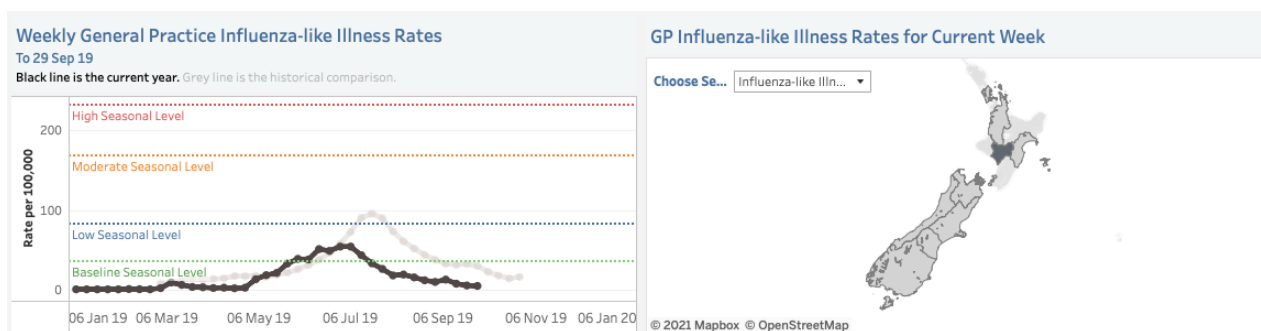[a]see `https://www.who.int/influenza`

Fig. 1.   Dashboard to visualise ILI Surveillance in New Zealand

Influenza pandemic (as seen with the ongoing COVID-19 pandemic). In an attempt to be more prepared globally, the WHO released a strategy to enhance Influenza surveillance worldwide. Part of this enhancement include: collecting and reporting better-quality epidemiological surveillance data. Member states develop and maintain the detection, assessment and reporting of the disease to the WHO within 48 hours building infrastructure for early warning signs for outbreaks of new virus strains or sub types.Several studies have demonstrated that advances in machine learning especially now that there is substantial historical data available, along with reliable data collection, can enable us to make improvements to disease surveillance and forecasting.[1-3]

In this paper, we apply machine learning models to predict ILI case counts from historic data. This helps to provide early warning to healthcare agencies which enables them to put in place preventative measures such as policy changes, increased healthcare capacity and raise public awareness. An advantage of using machine learning algorithms for the time series forecasting problem presented in this paper is that they can make predictions based on the ground truth and support the use of larger feature space. These could be past observations for single or multiple time series variables such as weather data which could impact the number of ILI cases observed in a community. Other general benefits of machine learning algorithms for time series forecasting over classical methods include the ability to support noisy features and support complex dependencies between features.

We propose a new time series forecasting algorithm called REACH (Randomized Ensembles of Auto-regression chains), which exploits dependencies in the predicted time intervals , allowing it to predict further into the future while maintaining high prediction accuracy. REACH is based on the concept of multi-label and multi-target classification, namely Ensembles of Classifier Chains.[4] So far, this concept has not been applied to a time series prediction problem. In this paper, we demonstrate the performance of this model in comparison to other standard models specifically for the prediction of ILI cases. The main contributions in this paper include: 1) a novel time series forecasting method, Randomized Ensembles of Auto-regression chains (REACH). 2) Improved prediction accuracy for ILI cases. 3) Predictions over a forecast horizon (i.e. the number of days we want to make predictions for) of 7 days,i.e. prediction of the number of ILI cases more accurately as we predict further into the future 4) An extensive evaluation on the New Zealand ILI cases from 2015 to 2018.

2. Related Work

Influenza surveillance is carried out in many countries around the world to help manage the health and economic impact of the disease. From data collections to creating reports takes time, which results in a delay in the reports getting to the authorities. This delay prevents authorities from being prepared in a timely manner, one way to get ahead of this delay is to forecast ILI activity for the coming week.

Several studies mentioned below have been conducted applying various novel and existing methods using various sources of data (for e.g. country specific ILI case counts, google search query data) . Epidemiological models such as SIR (Susceptible Infectious Removed) and SEIR (Susceptible Exposed Infectious Removed) models are commonly used for respiratory disease forecasting.[5] These models provide a valuable tool for management and control of an infectious disease, however, such models are pathogen specific and require a good understanding of the underlying disease dynamics. This is, in contrast to, statistical and machine learning models (including the proposed method REACH) which uses the ground truth to generate forecasts of syndromic ILI (which can be caused by several different respiratory pathogens). Because they forecast something entirely different to machine learning and statistical models such models are not used for comparison to the results in this study.

Auto-regression based statistical methods such as ARIMA (Autoregressive Integrated Moving Average)[6] have been a popular choice in time series forecasting. A study to forecast monthly incidence of Influenza in China[7] used an ARIMA model to forecast cases and proposed a new method to forecast mutations in the influenza A virus. Analysis using multiple sources of data over fifteen Latin American countries by Chakraborty *et. al.* implemented statistical methods as a baseline.[8] Zhang *et. al.* constructed a SARIMA (Seasonal Autoregressive Integrated moving average) model using Australian Influenza surveillance data and local search query data to predict seasonal influenza epidemics.[9] A method by Yang *et.al.* called ARGO (Autoregression with google search) base on linear regression using google search queries time series and historical ILI has demonstrated good results for ILI forecasting in the United states.[10]

In the last two decades several studies have demonstrated the use of machine learning approaches generating comparatively better forecasts.[11,12] A study by Santillana *et. al.* used three machine learning methods for forecasting, namely stacked linear regression, support vector machines (linear and using an independent feature mapping called radial basis function kernel (RBF)), and Regression trees with Adaboost.[13] More recently, Deep learning methods based on convolution and recurrent neural networks demonstrated good forecast accuracy when applied to large-scale data.[14,15] The use of LSTM(long short term memory) and gated recurrent neural networks are now being applied to time series modelling, and now to forecast infectious diseases. For example, a study published in 2020 by Chimmula *et.al.* used LSTM to forecast COVID-19 transmission in Canada.[16] In a recent study Wu *et. al.* have developed a transformer based model to forecast Influenza in the United stated which performed better when compared to ARIMA, LSTM and Seq2Seq models. Along with the methodologies, the data source and quantity of data available contribute equally to more accurate ILI forecasting. Most studies discussed here especially ones applying deep learning models have been conducted

on data for larger (by population) countries (in comparison to the population of New Zealand) such as the United States, Japan, China, Latin America, Australia, etc. Methods to improve forecast accuracy can benefit the way in which disease surveillance is carried out in countries such as New Zealand where its size and population is small and so is the resulting clinical data.[17] The smaller data set limits the use of deep learning models to work as intended, hence, they will not be applied in this study.

3. Background

A time series is a sequence of quantitative observations at successive time steps.[6] Time series forecasting is a type of forecasting where the past observations of a variable are used to incorporate its underlying relationships and patterns over time. A model that best captures these underlying relationships is then used to predict future values of that variable. A given time series comprises three systematic and one non-systematic (or random) component. The systematic components include trend, seasonality and level. The models described below are used in this study to compare with the proposed method REACH.

One of the most well-known and widely used auto-regression based time-series models are ARIMA models.[6,18] ARIMA models contain three components:The autoregressive component, a differencing component and an error component. The above-mentioned components can all be used to optimize ARIMA models, but newer implementations allow for automatic parameter optimisation.[6]

ETS models are a univariate time series forecasting method with an underlying state space model which consists of an error term (E), a trend component(T), a seasonal component(S) and a level component. In these models the weight decays exponentially with time, i.e. current observations have larger weights compared to past observations..[6]

Prophet uses a decomposed time series with three major components, namely trend, seasonality and holidays. It is a regression model at its core, which is made interpretable due to the decomposition of the series which allows users to fine tune specific components of the time series.[1]

The TBATS model is a type of Exponential smoothing state space model which can be used when the data has multiple or complex seasonality. Each seasonality is modeled by a trigonometric representation as a means to decompose the complex seasonal time series.[19]

4. Methods

Multistep forecasting, unlike one-step forecasting, has several underlying challenges such as accumulation of errors, reduced accuracy and increased uncertainty.[20] Many strategies have been proposed to overcome the ambiguity of multistep forecasting such as the 'recursive' strategy, where one time series model is estimated and each new forecast is computed using previous forecasts. A commonly applied strategy to mitigate error accumulation in multistep forecasting is the 'direct' forecasting strategy where a separate model is developed to forecast each lead time. A direct-recursive hybrid strategy allows us to incorporate the benefits of both methods.[21]

4.1. Reach – *Randomized Ensembles of Auto-regression Chains*

We propose a novel direct-recursive hybrid approach to overcome the limitations of direct and recursive approaches. The novel algorithm, REACH, stands for randomized ensembles of auto-regression chains. This approach is based on Ensembles of Classifier Chains[4] which are used in both multi-label and multi-target predictions. REACH uses individual regressors for each day in the horizon, similar to a simple auto-regression, however, instead of simply predicting the horizon in the natural order (i.e. first predict day 1 into the future, then use the prediction of day 1 as additional input to predict day 2 into the future and so on) REACH randomly shuffles the order of the days in the horizon to be forecast.

This method is based on the hypothesis that in some cases it might be easier to predict days that are far in the future directly as compared to predicting all days in a naturally ordered chain as auto-regressors do. A standard auto-regression has two main disadvantages. Firstly, it accumulates and propagates errors over the horizon. While only small errors might be introduced on the initial days, over the entire horizon, the errors build up and predictions get worse as we move into the future. Secondly, reverse time relationships are ignored. While time moves forward in one direction, we might make more accurate predictions of the initial days if the predicted values of the later days are known, even if those values for the later days might contain a larger error.

The proposed method REACH addresses the drawbacks of auto-regression and implements ensembles of auto-regression chains and predicts each day of the forecast horizon in a random order for every chain. Once the order of days to predict in the horizon is randomised, one regression model is then trained for each day in the horizon. The first model simply takes the historical data as input to predict the first random day in the chain. The second model takes the historical data and the predictions of that first random day in the chain as input, and so on. This allows multiple relationships among days in the past and future to be captured. We build ensembles of these randomly ordered auto-regression chains to strengthen the effect. The number of chains in the ensemble is a variable that can be adjusted to optimise the results. For simplicity, we omit the ensembles in the pseudo code. Overall, in a single chain, each day in the horizon is predicted, the algorithm is repeated n times and the median of the ensemble is used as the final prediction. The training of REACH is given in detail in Algorithm 1.

To the best of our knowledge, classifier or regression chains have not been used in the context of time series forecasting so far and REACH has the potential to improve time series forecast beyond the prediction of ILI cases.

5. Data

The regression models used in this study are applied to 2 data sets; namely, weather data/meteorological data and Sentinel ILI data of the Auckland region of New Zealand starting from the year 2015 until 2018.

The weather data is obtained from a public national climate database hosted by The National Institute of Water and Atmospheric research (NIWA) through the CliFlo [b] interface.

---

[b]https://niwa.co.nz/information-services/cliflo

Input: forecast window size (horizon) h, input size (lag) l, time series T of length n
Output: trained models $f_1, \ldots, f_h$
```
// generate a random order
```
$\omega = shuffle(1, \ldots, h)$
```
// calculate number of instances
```
$m = n - (l + h) + 1$
```
// store visited horizon
```
$x = \{\}$
```
// for each entry in the horizon
```
foreach $i \in 1, \ldots, h$ **do**
$\quad x = x \cup \omega[i]$
```
    // initialize training data
```
$\quad D = []_{m \times l + i}$
$\quad$ **foreach** $j \in 1, \ldots m$ **do**
```
        // set j-th row
```
$\quad\quad D_j = T[j, \ldots, (j+l), (j+l+x[1]), \ldots, (j+l+x[i])]$
$\quad$ **end**
```
    // train model with last column as target
```
$\quad f_i = train(D)$
end

Algorithm 1: Training of the randomized chain regressors of REACH

Sentinel ILI data is collected via participating Sentinel GP's who make a record of patients showing symptoms that fit the definition of ILI prescribed by the World Health Organisation. These data are reported during the winter surveillance period when there usually is ILI activity, i.e. weeks 18-39. The ILI data used in this study is public health information of New Zealand which cannot be shared due to ethics and privacy requirements.

The final multivariate data set used in this study consists of these features:Total number of reported cases per day, Maximum Temperature, Minimum Temperature, Rainfall, Relative humidity and Dew point.

ILI has some seasonal patterns which is reflected in the data. Figure 2 shows the distribution of a 10 day moving average of the daily data for the number of daily ILI cases reported by Sentinel GPs in New Zealand. The lack of ILI counts during the summer months is due to no surveillance activity and to take this into account in our data set the counts for these months are set to 0. There are 6 missing ILI case count values during the surveillance season in the entire data set and these are imputed using the average number of cases from the day before and the day after.
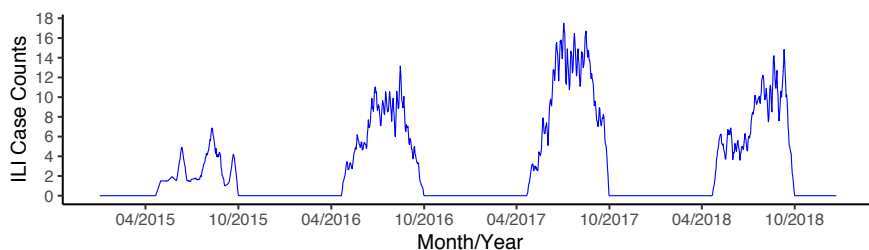

Fig. 2.   Distribution of the 10 day moving average of ILI case counts as seen in the GP ILI Data

To implement machine learning approaches for time series predictions, the time series problem is converted into a supervised learning problem. Creating the supervised learning problem from a univariate or multivariate time series requires creating *lag* (i.e. the values from n days in the past) values which can be treated as input features in order to make

predictions. *Lag* values are the values of the time series feature from a previous time step which can be used as input variables to forecast future time steps. In the univariate time series, the feature space is populated using the lag variables of the time series feature that is being forecast (target variable). In a multivariate time series (i.e. GP and Weather data), the features include lag observations from the target and additional variables.

6. Experiments

In this study, we want to compare the performance of traditional statistical methods, machine learning models and the application of the same machine learning models using the proposed REACH method. We used the following regression models in our experiments to forecast ILI case counts: (1) Autoregression (2) ARIMA (3) ETS (4) TBATS (5) Prophet (6) Random Forest model - recursive approach (7) Random Forest model applied using REACH.

All the models compared in this study are implemented so as to take into account the ILI count observations of the past 7 days (i.e. 7 *lags*) as input features immediately preceding the day for which the forecast is being made. The current surveillance method in New Zealand has a one week lag between gathering current case counts and getting the information to policy makers. Hence the experiments are setup to make predictions for 7 days in the future.

Auto regression, ARIMA, ETS and TBATS models were run using from the forecast package in R. The number of trees selected in the Random Forest[22] model was 500. The optimum number of variables to sample at random in each tree was calculated based on the lowest OOB (Out of bag) error for each horizon before training the model. To implement Prophet, the Trend change points parameters were set to the range of 0.001-0.5 as recommended to allow for flexible trends. Bayesian sampling of 500 MCMC samples per chain with 4 chains for posterior inference are applied to generate uncertainty intervals for trend and seasonality. All the non deterministic models, i.e. random forest, REACH and Prophet were repeated 10 times to observe variance in the accuracy of the forecasts. Each model is applied to both the univariate and the multivariate time series with the exception of ETS and TBATS because they do not allow for the use of external regressors (i.e. they are univariate forecasting methods).

As is common practice in time series forecasting,[6] the test and training sets are selected sequentially, so as to not miss any overarching seasonality and trend that might emerge in the historical data. This also helps to keep experimental runs as close as possible to the real-world application, where the model will be used to make future predictions based on the past and present data. The data used in this study were from year 2015 to year 2018 from the Sentinel GP source. We used data from year 2015 to year 2017 for training the model, and data from year 2018 for testing the model.

The Models used in this study are evaluated using a variation of leave-one-out cross validation.The results from the cross-validation can be found here `https://github.com/nooriyahp/PSB2022.git` . The method used for the cross validation of REACH is based on prequential cross validation evaluation of data streams[23] which addresses the time scale and the split between test and training over a sequential data set (i.e. the splits are not random). To implement this, the parameters of the models implemented were the same as the first experiment with the exception that in the first iteration the training set is small (i.e. the 7 days

Table 1.   RMSE for forecast horizons h 1 to 7 using univariate GP ILI time series

h	AR	ETS	ARIMA	TBATS	RF	REACH	Prophet
1	4.19	4.83	4.21	4.97	4.32 ±0.00	**4.18 ±0.01**	5.91±0.03
2	4.56	4.94	4.51	5.21	4.59 ±0.06	**4.39 ±0.01**	5.93±0.03
3	4.64	4.94	4.65	5.3	4.48 ±0.05	**4.46 ±0.02**	5.98±0.02
4	4.72	4.87	4.77	5.22	4.50 ±0.01	**4.45 ±0.02**	6.05±0.04
5	4.75	4.88	4.81	5.23	4.54 ±0.02	**4.49 ±0.01**	6.10±0.03
6	4.76	5.04	4.81	5.28	4.56 ±0.03	**4.52 ±0.01**	6.12±0.02
7	4.79	5.31	4.87	5.47	4.62 ±0.02	**4.59 ±0.01**	6.10±0.02

Table 2.   MAE for forecast horizons h 1 to 7 using univariate GP ILI time series

h	AR	ETS	ARIMA	TBATS	RF	REACH	Prophet
1	**3.21**	3.63	3.24	3.25	3.37 ±0.01	3.27 ±0.01	4.78±0.02
2	3.56	3.8	3.55	3.53	3.56 ±0.03	**3.50 ±0.02**	4.81±0.02
3	3.63	3.9	3.66	3.61	**3.57 ±0.02**	**3.57 ±0.02**	4.84±0.02
4	3.66	3.96	3.70	3.67	3.58 ±0.01	**3.57 ±0.02**	4.89±0.03
5	3.69	3.96	3.75	3.69	**3.55 ±0.01**	3.60 ±0.01	4.94±0.02
6	3.69	3.93	3.74	3.68	**3.60 ±0.01**	**3.60 ±0.01**	4.96±0.02
7	3.72	4.09	3.78	3.73	3.63 ±0.02	**3.61 ±0.00**	4.93±0.02

lag only), a model was trained for each time step in the series. The training set gets larger as the number of iterations increase. Hence we train almost as many models as the days in the time series. Performance for both experiments (i.e. train-test split and cross-validation) were calculated for the predictions of the year 2018.

7. Results

The performance metrics used are root mean square error (RMSE) and Mean Absolute error (MAE) which are both a measure of difference between the predicted and actual values. Tables 1 and 2 show the RMSE and MAE for each day in the horizon for both the univariate time series forecasts. The experiments show that REACH consistently performs well over the 7 day horizon compared to the other approaches. Specifically on forecast horizon 7, REACH has the highest prediction accuracy on the last day of the forecast horizon, hence demonstrating that the proposed method help maintain prediction accuracy over a longer horizon. Note that we give variances only for Random Forest, REACH and Prophet since ARIMA, Autoregression, ETS and TBATS are deterministic.

The results also show a rather small variance in the accuracy of REACH compared to other algorithms. This is caused by using ensembles that return the median as the final prediction which moves the results closer together. The Prophet model has the highest error in predictions as compared to the other models. This could be due to the fact that Prophet was developed mainly for non-statisticians to easily use time series forecasting specifically in "business use cases" mainly at Facebook.[1] There is an increasing number of studies that are now using Prophet models, mostly for business and economics time series data. Studies comparing ARIMA and Random Forest models in Influenza surveillance demonstrate that Random Forest models provide enhanced prediction ability.[18,24] We draw a similar conclusion, i.e., prediction accuracy improves when using Random Forest models to forecast ILI.

Tables 3a and 3b are the RMSE and MAE for each day in the horizon for the multivariate time series forecasts. The comparison of performance between models shows that REACH

performs better than other models consistently throughout the forecast horizon. The RMSE and MAE for both sets of experiments calculate the difference in actual and predicted values slightly differently. While the RMSE values consistently shows lowest errors for the proposed method, the MAE value show lower or equal errors in the shorter horizons for Autoregression, Random forest and ARIMA for the univariate and multivariate analysis respectively. It is worth noting that for the day 7 forecast REACH shows the lowest error in both metrics. Comparison of performance metrics of the univariate and multivariate time series show that adding another source of relevant data (in this case the weather data) generally improves performance of all the models including REACH.

Figure 3a plots the change in RMSE over the forecast horizon when applied to univariate time series data. For forecast horizon of 1, we see the REACH has the lowest RMSE.

(a) Univariate (GP ILI) (b) Multivariate (GP ILI and Weather)
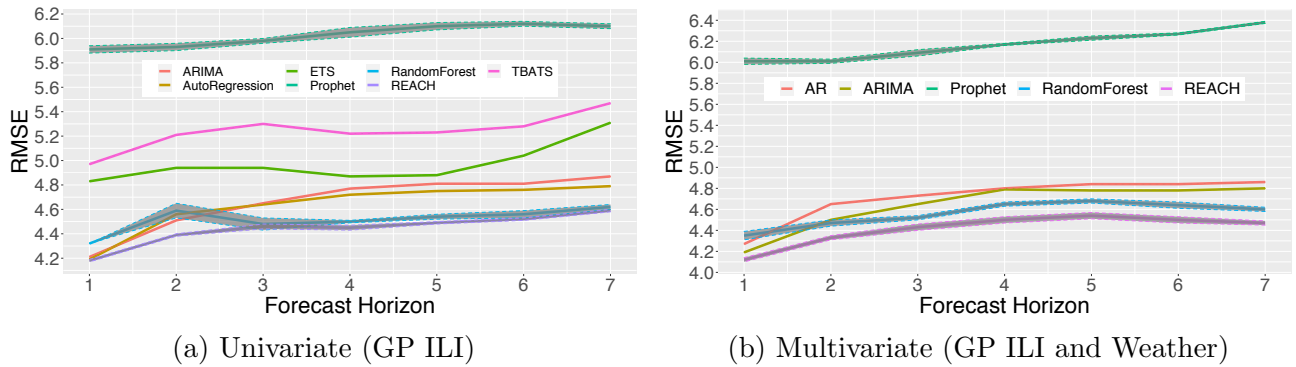
Fig. 3. Plots comparing RMSE's of the models applied to the univariate and multivariate data sets compared over each day of the forecast horizon. The grey shaded area shows the variance in error observed.

Figure 3b illustrates the change in RMSE over the forecast horizon for multivariate time series data. For each day in the forecast horizon (Days 1 to 7), our experiment demonstrates the predictions made by the REACH model report the lowest error. Note that in this case, REACH consistently achieves a lower RMSE than other methods, which demonstrates that it is better at using the additional input features. ARIMA and other models might not be able to cope with a large feature space while the Random Forests benefit from more features.

Due to space constraints, the results from the cross-validation experiments are available

Table 3. RMSE MAE for forecast horizons h 1 to 7 using multivariate GP ILI and Weather time series

h	AR	ARIMA	RF	REACH	Prophet	h	AR	ARIMA	RF	REACH	Prophet
1	4.27	4.19	4.35 ±0.04	**4.12 ±0.02**	6.01±0.03	1	3.25	**3.20**	3.45 ±0.03	3.26 ±0.01	4.83±0.01
2	4.65	4.50	4.47 ±0.03	**4.33 ±0.02**	6.01±0.02	2	3.58	**3.39**	3.55 ±0.02	3.45 ±0.01	4.84±0.02
3	4.73	4.65	4.52 ±0.02	**4.43 ±0.03**	6.09±0.03	3	3.64	3.56	3.61 ±0.02	**3.54 ±0.01**	4.93±0.03
4	4.80	4.79	4.65 ±0.02	**4.5 ±0.03**	6.17±0.01	4	3.68	3.66	3.66 ±0.01	**3.61 ±0.01**	5.02±0.03
5	4.84	4.78	4.68 ±0.02	**4.54 ±0.03**	6.23±0.02	5	3.71	3.68	3.70 ±0.02	**3.61 ±0.02**	5.06±0.01
6	4.84	4.78	4.64 ±0.03	**4.50 ±0.03**	6.27±0.01	6	3.72	3.68	3.64 ±0.03	**3.57 ±0.01**	5.08±0.03
7	4.86	4.80	4.60 ±0.02	**4.47 ±0.02**	6.38±0.01	7	3.72	3.69	3.65 ±0.02	**3.55 ±0.01**	5.15±0.02

(a) RMSE (b) MAE

here https://github.com/nooriyahp/PSB2022.The cross validation results show a higher prediction accuracy which is as expected because the models in this experiment were trained on a larger part of the data set. The cross validation results are also reassuring that our models were not overfitting.

From these results, we conclude that the proposed method is a data-driven approach to forecasting ILI counts and predictions can be enhanced even further by adding more relevant features. Future work will involve implementing the REACH method on larger data sets that contain supporting information to help predict ILI case counts such as twitter data and google search queries. In the future work we will apply models like ARGO (see section 2) which use google query data to forecast ILI for a comparison with REACH.

The decrease in the error for both Random Forests and REACH for day 6 and 7 initially might be counter-intuitive. However, we want to point out that the change is rather small. In the prediction of ILI cases, to predict the case numbers in a week seems reasonable as there are external factors that may affect the weekly fluctuation of cases such as, the likelihood of patients visiting the GP on the weekend which might affect predictions in the 7 day cycle.

Statistical significance tests of the prediction accuracy of each model compared with REACH was performed using the Diebold-Mariano method.[25] The results in Table 4 confirms that the difference in the error metrics for each model compared with REACH are statistically significant (i.e. P-values <0.05).

Table 4. P-values of the Diebold-Mariano test for statistical significance of the predictions made by each algorithm compared to the proposed REACH algorithm on the GP only data set and GP with Weather data set

$Models$	GP(P-value)	$WeatherGP$(P-value)
REACH vs Autoregression	0.0013	$5.91e^-05$
REACH vs ETS	$3.07e^-05$	n/a
REACH vs ARIMA	0.0022	0.00057
REACH vs TBATS	$4.41e^-08$	n/a
REACH vs RF	0.024	$5.3e^-05$
REACH vs PROPHET	$6e^-10$	$1.4e^-08$

8. Conclusion

Predicting ILI case count information ahead of time has the potential to substantially reduce the impact of the flu season, as it allows time for public health interventions. In this paper, we propose a novel method REACH and demonstrate that it performs better than state-of-the-art methods for a forecast horizon of 7 days. REACH employs a randomized chain of ensembles of models enhancing the prediction ability of the Random Forest model even further for predicting a long forecast horizon (see Figure 3a and 3b). We evaluated the performance of the model on ILI case counts in New Zealand from 2015 to 2018. The results presented here demonstrate that the proposed method improves prediction accuracy for ILI cases. The extensive evaluations support these findings by the low errors of the predictions generated by REACH.

While additional data such as Weather data used in this study improves the prediction accuracy of the proposed method for most days of the forecast horizon. There are some fundamental differences in the way multivariate time series are implemented in all the models. All other information outside of the primary time series (GP ILI case counts) for which predictions are generated, are passed into the statistical model (i.e. ARIMA and Prophet) as external regressors. All models used for the multivariate time series forecasting are trained using the additional time series data (i.e. weather data). A fundamental difference lies in the forecasting step of the ARIMA and Prophet models the forecasts or known future values of the external regressors are required for the prediction of GP ILI counts. Unlike Prophet and ARIMA, at the forecasting step, REACH uses only past values of the weather variables and yet generates better forecasts for the future values of GP ILI case counts. This fundamental difference in the implementation of a multivariate time series to the models used in this study is an advantage of the proposed REACH method while using additional features.

The use of private health data sets is quite limiting in terms of making the data public due to ethical and privacy restrictions. Another limitation of the REACH algorithm requires training multiple models as compared to autoregression, ARIMA and Prophet models which only require to be trained once to forecast all forecast horizons. The effect of this limitation can be reduced by using models that are fast to run, such as random forests.

9. Acknowledgements

References

1. S. J. Taylor and B. Letham, Forecasting at scale, *The American Statistician* **72**, 37 (2018).
2. V. Papastefanopoulos, P. Linardatos and S. Kotsiantis, Covid-19: A comparison of time series methods to forecast percentage of active cases per population, *Applied Sciences* **10**, p. 3880 (2020).
3. J. R. Ortiz, V. Sotomayor, O. C. Uez, O. Oliva, D. Bettels, M. McCarron, J. S. Bresee and A. W. Mounts, Strategy to enhance influenza surveillance worldwide, *Emerging infectious diseases* **15**, p. 1271 (2009).
4. J. Read, B. Pfahringer, G. Holmes and E. Frank, Classifier chains for multi-label classification, *Machine learning* **85**, 333 (2011).
5. R. N. Thompson, Epidemiological models are important tools for guiding covid-19 interventions, *BMC Medicine* **18**, 1 (2020).
6. R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice* (OTexts, 2018).
7. X. Song, J. Xiao, J. Deng, Q. Kang, Y. Zhang and J. Xu, Time series analysis of influenza incidence in chinese provinces from 2004 to 2011, *Medicine* **95** (2016).
8. P. Chakraborty, P. Khadivi, B. Lewis, A. Mahendiran, J. Chen, P. Butler, E. O. Nsoesie, S. R. Mekaru, J. S. Brownstein, M. V. Marathe *et al.*, Forecasting a moving target: Ensemble models

for ili case count predictions, in *Proceedings of the 2014 SIAM international conference on data mining*, 2014.

9. Y. Zhang, L. Yakob, M. B. Bonsall and W. Hu, Predicting seasonal influenza epidemics using cross-hemisphere influenza surveillance data and local internet query data, *Scientific reports* **9**, 1 (2019).

10. S. Yang, M. Santillana and S. C. Kou, Accurate estimation of influenza epidemics using google search data via argo, *Proceedings of the National Academy of Sciences* **112**, 14473 (2015).

11. A. K. Palit and D. Popovic, *Computational intelligence in time series forecasting: theory and engineering applications* (Springer Science & Business Media, 2006).

12. N. K. Ahmed, A. F. Atiya, N. E. Gayar and H. El-Shishiny, An empirical comparison of machine learning models for time series forecasting, *Econometric Reviews* **29**, 594 (2010).

13. M. Santillana, A. T. Nguyen, M. Dredze, M. J. Paul, E. O. Nsoesie and J. S. Brownstein, Combining search, social media, and traditional data sources to improve influenza surveillance, *PLoS Comput Biol* **11**, p. e1004513 (2015).

14. N. Wu, B. Green, X. Ben and S. O'Banion, Deep transformer models for time series forecasting: The influenza prevalence case (2020).

15. K. Chen, G. Chen, D. Xu, L. Zhang, Y. Huang and A. Knoll, Nast: Non-autoregressive spatial-temporal transformer for time series forecasting, *arXiv preprint arXiv:2102.05624* (2021).

16. V. K. R. Chimmula and L. Zhang, Time series forecasting of covid-19 transmission in canada using lstm networks, *Chaos, Solitons & Fractals* **135**, p. 109864 (2020).

17. L. Lopez, T. Wood, N. Prasad and S. Huang, Influenza surveillance in new zealand 2016 (2011).

18. M. J. Kane, N. Price, M. Scotch and P. Rabinowitz, Comparison of arima and random forest time series models for prediction of avian influenza h5n1 outbreaks, *BMC bioinformatics* **15**, p. 276 (2014).

19. A. M. D. Livera, R. J. Hyndman and R. D. Snyder, Forecasting time series with complex seasonal patterns using exponential smoothing, *Journal of the American Statistical Association* **106**, 1513 (2011).

20. S. B. Taieb and G. Bontempi, Recursive multi-step time series forecasting by perturbing data, in *2011 IEEE 11th International Conference on Data Mining*, 2011.

21. G. Bontempi, S. B. Taieb and Y.-A. Le Borgne, Machine learning strategies for time series forecasting, in *European business intelligence summer school*, 2012.

22. A. Liaw, M. Wiener *et al.*, Classification and regression by randomforest, *R news* **2**, 18 (2002).

23. A. Bifet, G. de Francisci Morales, J. Read, G. Holmes and B. Pfahringer, Efficient online evaluation of big data stream classifiers, in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15 (Association for Computing Machinery, New York, NY, USA, 2015).

24. T. Petukhova, D. Ojkic, B. McEwen, R. Deardon and Z. Poljak, Assessment of autoregressive integrated moving average (arima), generalized linear autoregressive moving average (glarma), and random forest (rf) time series regression models for predicting influenza a virus frequency in swine in ontario, canada, *PloS one* **13**, p. e0198313 (2018).

25. F. X. Diebold and R. S. Mariano, Comparing predictive accuracy, *Journal of Business & economic statistics* **20**, 134 (2002).