Journal of Coastal Life Medicine
www.jclmm.com
ISSN: 2309-5288(Print)/2309-6152(Online)
Volume 6 No.1 (2018), Page No. 01 – 05
Article History: Received: 05 January 2018, Revised: 12 February 2018, Accepted: 15 February 2018, Publication: 31 March 2018

The Effectiveness of Disease Prediction in Enhancing Patient Satisfaction at the Community Level

Musa Olasunkanmi

Department of Zoology, Ahmadu Bello University, Nigeria

Abstract

In contemporary society, trends in infectious disease occurrences, incidence, and prevalence are unknown in the majority of cases, complicating efforts geared towards disease prediction. In this study, the motivation was to contribute to infectious disease prediction through deep learning algorithms' parameters' optimization, with big data in the healthcare industry on consideration. Specifically, there was a comparison between the performances of the long-short term memory (LSTM) learning algorithm and deep neural network model with the performance of ARIMA (autoregressive integrated moving average) algorithm. Three infectious diseases were examined to discern model reliability and validity in under different experimental conditions. From the findings, this study established that the performance of LSTM and DNN frameworks is superior to that of ARIMA. Relative to chickenpox prediction, there was improvement by 19% and 25% after implementing LSTM and top-10 DNN models, respectively. Also, LSTM exhibited better accuracy while DNN exhibited more performance stability, especially with the spread of infectious diseases. The implication for the healthcare industry is that this study's findings could be used to inform some of the ways in which reporting delays could be eliminated or minimized, improving on the current surveillance systems.

Introduction

When pathogens are from another animal or persons, they could cause infectious diseases to another individual [1]. The harm stretches beyond the individual level to affect aspects operating on a macro scale. The implication is that infectious diseases are a critical social issue [2]. Recent studies point to different countries' engagement in a comprehensive process of surveillance for infectious disease outbreaks, with a continuous and systematic interpretation, analysis, and collection of vectors [2-4].

Regarding the case of conventional reporting system, certain reports from medical firms (regarding infectious disease) tend to be incomplete, an adversity that attracts reporting system delays [5]. A specific example is that in which traditional surveillance systems for influenza have witnessed two-week delays between report preparation stages and report dissemination [6, 7]. With delayed and occasionally missing reports, the trickle-down and negative effect is that early interventions aimed at curbing infectious diseases could be hindered [8, 9]. Therefore, the need to establish relevant data-based prediction models that focus on infectious diseases could not be overstated [10]. This study sought to implement and evaluate the performance of a prediction model for infectious diseases, especially in real-time. Also, the study strived to give insight into a disease prediction model capable of understanding the degree to which trends in infectious disease occurs, paving the way for cost minimization at the societal level.

Methods

With infectious disease prediction being the central subject, the study focused on four data forms. They included humidity, temperature, social media big data, and search query data. The data that

Journal of Coastal Life Medicine
www.jclmm.com
ISSN: 2309-5288(Print)/2309-6152(Online)
Volume 6 No.1 (2018), Page No. 01 – 05
Article History: Received: 05 January 2018, Revised: 12 February 2018, Accepted: 15 February 2018, Publication: 31 March 2018

was used was that which had been documented between early 2016 and mid 2017. For the weather data concerning parameters of humidity and temperature, the information was gained from a regional meteorological department. Average humidity was taken in terms of percentage while average temperature was taken in the form of degree Celsius. With the Python Selenium library also used for a web crawler, Twitter was the selected site. This site was the platform from which there was the collection of infectious disease social media big data. Focusing was on tweets that had mentioned infectious disease on a daily basis. It is also notable that an infectious disease web statistics platform was used for collecting infectious disease data.

For the designed surveillance model, it relied on weather data, twitter data, and non-clinical search data. Indeed, disease occurrence was the output variable while optimal variable combination reflected the input variables. There was also a division of the OLS dataset to obtain a test data subset and a training data subset. The ratio was 2:8.

Results

There was a seven-day lag application to the respective datasets for infectious disease. For the collected data, the number of days considered were 569 in the regression model. The OLS results that were obtained were summarized in a tabular form as shown in the table below.

Disease	R^2	Adjusted R^2	F	p	Variable	Coefficient	T	p
Chickenpox	0.4077	0.4035	97.0659	<0.001	Naver	4.4569	18.2096	<0.001
					Twitter	0.2162	0.9769	0.3290
					Temperature	−3.8421	−8.7717	<0.001
					Humidity	−0.8919	−3.1027	0.0020
					Intercept	121.9521	6.6302	<0.001
Scarlet fever	0.2867	0.2817	56.6851	<0.001	Naver	2.1956	12.6929	<0.001
					Twitter	−1.981	−1.3940	0.1639
					Temperature	0.2559	1.5660	0.1179
					Humidity	−0.5369	−4.4766	0.0491
					Intercept	68.5623	9.4170	<0.001
Malaria	0.3863	0.3819	88.7462	<0.001	Naver	0.0649	7.1140	<0.001
					Twitter	0.0369	1.6594	0.0976
					Temperature	0.0770	5.7623	<0.001
					Humidity	0.0129	1.6421	0.1011
					Intercept	−1.8257	−3.7883	<0.001

For the humidity, temperature, and Naver search queries, significant results were reported for the scarlet fever model when the chickenpox model was implemented as an infectious disease regression framework. For the case of the malaria framework, significant results were found in relation to temperature and the Naver search queries. It is also notable that for all the three infectious diseases, the Naver search query data proved significant. However, for all the three infectious diseases, Twitter data did not prove significant. Therefore, previous studies had documented that Internet search query data is insightful relative to infectious disease prediction model design, but this study did not establish any significant role that Twitter data could play relative to the same aspect of model design. However, a unique finding that was reported was that

Journal of Coastal Life Medicine
www.jclmm.com
ISSN: 2309-5288(Print)/2309-6152(Online)
Volume 6 No.1 (2018), Page No. 01 – 05
Article History: Received: 05 January 2018, Revised: 12 February 2018, Accepted: 15 February 2018, Publication: 31 March 2018

the Twitter data could be used towards identifying a framework exhibiting the highest adjusted R-squared value.

For the parameter of temperature, this study established that it exhibited a strong correlation with the selected infectious diseases – apart from malaria. From the coefficients' values, it can be seen that the Naver search query data reflected scarlet fever and chickenpox's most significant variables. The values stood at 2.1956 and 4.4589, respectively. At 0.0770, the temperature values were found to be the most significant variables for the case of malaria. Given that for all the three infectious diseases, Naver search query data remained significant, it proved important and suitable in relation to infectious disease prediction.

Similar to the case of OLS, there was the evaluation of the seasonal ARIMA framework via the same data. At this stage, focus was on the model's behavior or performance based on the parameters of RMSE and AIC.

The next phase involved a focus on LSTM and DNN prediction frameworks. In the respective order, values in the parentheses for each of the two models represented the optimizers, activation functions, and numbers of epochs that the models used. In situations where lower RMSE values were obtained, these aspects were representative of smaller differences between the predicted and actual values, pointing to higher prediction performance of the given model.

Further investigation involved the process of establishing an optimal performance model for the three infectious diseases. To achieve this objective, which was spearheaded in a healthcare environment marked by big data, each model was applied to each disease and the three sets of results reported. Indeed, in this study, compared to traditional ARIMA approaches, outstanding performance was reported for the case of the deep learning framework. With the LSTM and DNN prediction algorithms for chickenpox considered in the entirety, the lowest RMSE was achieved for the optimal models at 27.33% and 27.22% and outperformed the ARIMA framework, respectively. For the LSTM and top 10 DNN frameworks, there was improvement in performance by average values of 18.78% and 24.45%, respectively. For the LSTM and DNN prediction frameworks, when scarlet fever was considered, their lowest RMSE values stood at 23.79% and 26.25% performance improvements, respectively; compared to the case of ARIMA models.

It is also worth noting that in this investigation, regarding the ARIMA framework that was employed, its effectiveness was confirmed in situations where infectious diseases' number of incidences was regular, without experiencing decreasing or increasing trends. However, it is imperative to note that most of the studies avow that when the actual data is considered, it could be irregular and also have trends. As such, the study demonstrated that when deep learning is employed, it forms a superior analytical framework through which data could be analyzed and, in turn, provide room for future situations' predictions. From some of the previous studies' scholarly reports, it has been established that the deep learning algorithm is better placed relative to the ability to ensure that any decreasing or increasing trends are followed in a manner that proves sufficient [7-10]. Additionally, this study demonstrated that the LSTM and DNN frameworks remain sensitive to increasing trends and decreasing trends, respectively.

Journal of Coastal Life Medicine
www.jclmm.com
ISSN: 2309-5288(Print)/2309-6152(Online)
Volume 6 No.1 (2018), Page No. 01 – 05
Article History: Received: 05 January 2018, Revised: 12 February 2018, Accepted: 15 February 2018, Publication: 31 March 2018

Conclusion

In summary, one of the critical social problems facing contemporary society entails infectious disease incidence and prevalence. Apart from causing damage at individual levels, infectious disease can be seen to yield widespread harm. Therefore, there has been growing research attention relative to the need to reduce social losses with which infectious diseases are associated. One of the specific trends that have been embraced is the prediction of diseases. In this experimental study, the objective lay in the establishment of a prediction framework for infectious diseases, especially one that would yield superior results compared to previously proposed models. To realize the latter specific objective, the designed framework employed deep learning techniques and various input variables. From a methodological perspective, there was the setting of optimal parameters based on the OLS-based technique of selecting variables. With optimal parameters, there was the performance of analyses of LSTM, DNN, and ARIMA.

Based on the optimal parameter usage to analyze OLS, findings demonstrated that for the respective infectious diseases, the regression frameworks exhibit significant outcomes. For example, considering the four variables, this study found that relative to the three infectious diseases in the entirety, the Naver search frequency demonstrated a significant correlation. To discern deep learning frameworks, this study also focused on ARIMA analysis, as well as OLS performance. Relative to the LSTM and DNN results, it was noted that both deep learning methods exhibit better predictions regarding the three infectious disease trends, compared to ARIMA and OLS frameworks. In addition, it was found that on average, the best performance comes with the implementation of DNN models. However, whereas the best performance arose from the implementation of the DNN frameworks, more accurate predictions arose from the implementation of the LSTM frameworks, especially in situations where there was a spread in infectious diseases. It is also notable that when the case of malaria was evaluated, an aspect that is worth acknowledging is that the disease's occurrences were fewer than the other two infectious conditions, an attribute that might have flawed the comprehensiveness and accuracy or reliability of the predictions. In the future, it becomes imperative to address this gap by conducting a similar investigation in scenarios where there are more occurrences of malaria.

Overall, this study was insightful because it revealed certain unique features with which the LSTM and DNN frameworks are associated. For instance, it was established that compared to the LSTM framework, the DNN framework yields smaller values relative to infectious disease prediction. Thus, if the intention is to project the disease occurrence's minimum value, the DNN framework is worth employing. However, if the aim of the prediction lies in maximum value prediction in relation to disease occurrences, the LSTM model is worth employing.

References

[1] Towers, S.; Afzal, S.; Bernal, G.; Bliss, N.; Brown, S.; Espinoza, B.; Jackson, J.; Judson-Garcia, J.; Khan, M.; Lin, M.; et al. Mass Media and the Contagion of Fear: The Case of Ebola in America. *PLoS ONE* 2015, *10*, e0129179, doi:10.1371/journal.pone.0129179.

[2] Huang, D.C.; Wang, J.F. Monitoring hand, foot and mouth disease by combining search engine query data and meteorological factors. *Sci. Total Environ*. 2018, *612*, 1293–1299, doi:10.1016/j.scitotenv.2017.09.017.

ISSN: 2309-5288(Print)/2309-6152(Online)
Volume 6 No.1 (2018), Page No. 01 – 05
Article History: Received: 05 January 2018, Revised: 12 February 2018, Accepted: 15 February 2018, Publication: 31 March 2018

[3] Tenkanen, H.; di Minin, E.; Heikinheimo, V.; Hausmann, A.; Herbst, M.; Kajala, L.; Toivonen, T. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Sci*. *Rep*. 2017, *7*, 17615, doi:10.1038/s41598-017-18007-4.

[4] Reece, A.G.; Reagan, A.J.; Lix, K.L.M.; Dodds, P.S.; Danforth, C.M.; Langer, E.J. Forecasting the onset and course of mental illness with Twitter data. *Sci*. *Rep*. 2017, *7*, 13006, doi:10.1038/s41598-017-12961-9.

[5] Shin, S.; Seo, D.; An, J.; Kwak, H.; Kim, S.; Gwack, J.; Jo, M. High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea. *Sci*. *Rep*. 2016, *6*, 32920, doi:10.1038/srep32920.

[6] Thapen, N.; Simmie, D.; Hankin, C.; Gillard, J. DEFENDER: Detecting and Forecasting Epidemics Using Novel Data-Analytics for Enhanced Response. *PLoS ONE* 2016, *11*, e0155417, doi:10.1371/journal.pone.0155417.

[7] Allen, C.; Tsou, M.; Aslam, A.; Nagel, A.; Gawron, J. Applying GIS and Machine Learning Methods to Twitter Data for Multiscale Surveillance of Influenza. *PLoS ONE* 2016, *11*, e0157734, doi:10.1371/journal.pone.0157734.

[8] Volkova, S.; Ayton, E.; Porterfield, K.; Corley, C.D. Forecasting influenza-like illness dynamics for military populations using neural networks and social media. *PLoS ONE* 2017, *12*, e0188941, doi:10.1371/journal.pone.0188941.

[9] Simon, T.; Goldberg, A.; Aharonson-Daniel, L.; Leykin, D.; Adini, B. Twitter in the Cross Fire—The Use of Social Media in the Westgate Mall Terror Attack in Kenya. *PLoS ONE* 2014, *9*, e104136, doi:10.1371/journal.pone.0104136.

[10] Tafti, A.; Zotti, R.; Jank, W. Real-Time Diffusion of Information on Twitter and the Financial Markets. *PLoS ONE* 2016, *11*, e0159226, doi:10.1371/journal.pone.0159226