INFECTIOUS DISEASE PREDICTION BASED ON PATIENT RECORDS USING MACHINE LEARNING ALGORITHMS - A CASE STUDY

K. Sathesh Kumar
Assitant Professor
Department of Computer Science and Information Technology
Kalasalingam Academy of Research and Education,Krishnankoil,
k.satheshkumar@klu.ac.in

Abstract

An environmental change plays an important role to infection outbreaks. Some maladies can modify their dissemination based on climatic factors. An environmental factor such as temperature, humidity, precipitation plays a vital role to change the spread of the aliments .For example, Malaria and dengue produces their enormous outbreaks in rainy seasons. Because the rainy seasons create a positive environment for mosquito production. Machine learning is the emerging tool to forecast the disease occurrence. In this paper, the researchers discuss three aliments COVID-19, malaria and flu outbreaks based on seasons. As well as the researcher predicts which season influences which infection based on health records. And also discuss various machine learning algorithms for predicting the seasonal wise occurrence of infections using Electronic patient health records. And also forecast the winter season is more appropriate season for spreading more disease outbreaks when compared to other seasons. This information is very useful for health analyst to prevent the people earlier

Keywords: Environmental condition, Machine learning, Maladiesand Health records etc

1. Introduction

The modern world has plenty of development in all fields such as Industries, medical field, Agriculture etc. In the similar way, the people population has increased in infinite manner. Now, the world population reached 7.8 billion. Till 2050 it will be reached 9.7 million. Because of this population, the people faces several problems such as globalization, Urbanization, global warming. The countless population will create lots of unemployment. Industrial development is the only solution to solve the unemployment problem.

The population growth and Industrial development are the main component of climate change. And urbanization, globalization and global warming are also the major factors of environmental change.Atmospheric modification leads to uneven rainfall and the highest temperature.Consequently, as the environmental change, so does the rate of infectious disease. Some climatic variation produces new pathogens.

Now a day, the world stalled by COVID-19. Weather change is the supreme element of producing a new disease. Many of the infectious aliments can modify their outbreaks based on the climatic factors such as Malaria can increase their outbreaks in rainy seasons. As well as, dengue

highly fabricates their parasites in summer and monsoon seasons. Most of the diseases are released their outburst depends upon the weather change.

In this several deadly diseases are also there such as COVID-19, Malaria, Dengue and flu etc. Still now the people are suffered by COVID-19. Around the world, lots and lots of people died by COVID-19. But the researcher also face struggle for predict the correct medicine for COVID-19. Malaria is one of the deadly diseases. It mostly affects infants, children and women. It causes the severe effects in pregnant woman. Flu infects children. As well as it makes the dangerous impacts to children.

Machine learning is the powerful technique to forecast these types of deadly infectious aliments earlier. This emerging technique is very helpful for the researcher to classify the diseases in deadly and non-deadly category. Machine learning is used in various fields such as Health care, Medicine, Bank sector and Business Administration etc. Climate forecasting is one of the major benefits of expert systems. Expert system techniques can help health organizations prevent the spread of severe infectious disease outbreaks in a variety of ways.

This paper is structured as follows. Section 2 describes the literature review of seasonal infectious diseases. Section 2.1 represents the detailed information of data. Section 2.2 illustrates the technique of data-preprocessing. Section 2.3 presents the technology of feature selection methods. Section 2.4 gives the details of proposed method. Section 2.5 presents a discussion of our findings and section 3 concludes the research work.

2. Background study

In **(**Indhumathi et al., 2020) the author specifies three vital factors to spread the deadly aliments depend on climatic variation. The first factor is human. Human plays an important role to transmit an epidemic from one human to another. The second one is pathogen. This is the main reason to spread the disease to human. And the third one is circumstances. Circumstance is the primary element to create the surrounding for produce the parasites. Without these three factors the infections cannot produce their outbreaks.

Climate change (Willox et al., 2015, Altizer et al., 2013, Bouzid et al., 2014) has the potential to disrupt human health and outbreaks of many infectious illnesses are at risk. In (Chowdhury et al., 2018) the researcher describes the relationship between the warm environment, dew, precipitation and seasonal epidemics. It combines the five years of weather data and patient records. By using these data, the author forecast the aliments spread in earlier

Based on (Moriyama et al., 2020, Kudo et al., 2019, Marr et al., 2019, Lowen et al., 2014 & Chan et al., 2011) chill and dry weather, flu and COVID- 19 can expanded their outbursts. The survey (Bukhari et al., 2020) found that the chill and warm conditions are related to the spread of flu and COVID -19. Analysis (Manogaran et al., 2017) identifies the dengue spread is highly increased in summer and monsoon seasons.

The studies (Ostfeld et al., 2015 &Rodrigue-Arias et al., 2013) identifies that the prolonged heat climate is more suitable to lay out their pathogens. The extreme hot weather creates several deadly epidemics in unpredicted places. This survey identifies the climatic factors are the primary factors to generating the infections. According to (Kokaze et al., 2001) weather condition may impact chickenpox infection elements. As well as the spread of chicken pox, grow their count of pathogens in frost and vernal seasons.

To ensure that the machine learning approaches (Jennifer et al., 2021) to forecast the seasonal wise occurrence of influenza outbursts. By using the expert systems, the author identifies the flu expanded their outburst in winter seasons. According to (Juhyeon et al., 2021), the author describes the relationship between the infectious diseases and environmental condition.

The study (Shirsath et al., 2018)made a prototype for cardiovascular diagnosis using PCA-based attribute selection. Reducing data dimensionality, as per the researchers, increases the classifier's prediction accuracy while decreasing the computational cost of the forecast. It could be done with a feature selection method that produces a new collection of attributes from the primary data, or an attribute extraction technique that selects a subcategory of its relevant features of the dataset.

This paper (Shaukat et al.,2015) implements a Chi-squared attribute evaluator for extracting the significant features and applies the random forest classification algorithm toward make a model for cardiovascular disease identification. In which they rank the features according to their importance. Chi-squared test, then delete the lowest-ranking function one at a time, constructing and testing models at each level until the model's precision reaches a plateau. They found that a rate of 83.7 percent was the most accurate model.

The article (Dahiwade et.al., 2019)proposes a standard variable selection approach of stepwise backward selection using p-values have been commonly utilized in the medical literature. The study(Chen et al., 2017) suggests Elastic Net attribute extraction method. It's similar to a Lasso extension with the addition of an L1 and L2 penalty.The L2 norm's characteristics promote a clustered impact, allowing strongly correlated variables to be held within the representation or discarded together. In a similar way to the Lasso, it also performs embedded variable collection.

The survey (Sah1 et al., 2017)decide which variables were chosen and to test the method's output on the validation collection, the Lasso feature selection model with the highest AUC was used. The studies (Klepac et al., 2014) provide the new knowledge about handling the coronavirus situations. It provides, Expert system and massive inputs enhance that Corona circumstances and brings the study of stopping corona outcomes. The researcher, five classification models are utilized to found the pathogen's detection using supervised learning algorithms.

When correlated to another Data Science processes, the XGBoosting model is the most significant accurate model for recognizing virus families. It found 100% of Severe Acute Respiratory Syndrome and mosquito-borne flavivirus, 84.61 % of flu viruses and 75% Zaire

ebolavirus also Hepatitis. As well as discover 88-89% of COVID-19. This paper (Abdul Rahman., 2017) compares the effects of coronavirus with other epidemics and determines how the ailments modify the entire society along with the world economy.

The research (Lopez et al., 2014) presents a geographically weighted regression model for identifying the spread of influenza in Vellore.The flu prevalence and meteorological conditions such as precipitation and air velocity are significantly associated throughout this model. Furthermore, this lead to an adverse association of influenza outcomes with temperature and rainfall.The Pearson's correlation coefficient framework is used throughout the review (Chandy et al., 2013) to determine the association with vector-borne disease and environmental degradation.

The research (Alfred et.al, 2021) discusses machine learning (ML) techniques can be used to stop dangerous infectious disease outbreaks from spreading (e.g., COVID-19). Machine learning algorithms could be used to anticipate and identify the severe viral disease. The machine learning methods, datasets, and quality measures used mostly for variety of applications in detecting and diagnosing the lethal infectious disease were not discussed in most publications.

This research, on the other hand, presents a literature evaluation based on two primary approaches to limiting the spread of lethal disease outbreaks (e.g., prediction and detection). As a result, the goal of this research is to look into the current state of the art, obstacles, and future work in using machine learning to detect and identify severe epidemics in the two groups mentioned previously.

4. Conclusion

Environmental change is the most primary fact for disease pathogens to spread their outbreaks. Based on the seasons, the epidemics outburst is differ. Based on the background study COVID-19, Malaria and flu spread their outbreaks mostly in winter seasons. The winter season is favorite period for aliments. Because this season has heavy rain falls. Rain is the supreme factor for producing mosquitoes. Mosquito cause several deadly diseases. Next, summer is second suitable period for aliments. It provides warm environment. Some parasites are survived in hot environment. Machine learning is one of the leading technologies to predict the infectious diseases. Early disease prediction is very helpful for health analyst to protect the people from the epidemics in earlier. This is very useful for health analyst to prevent the people from the deadly infections earlier.

References

1. Indhumathi, K., & Sathesh Kumar, K. (2020). A review on prediction of seasonal diseases based on climate change using big data. *Materials Today: Proceedings, Elsevier*,pp.1-5.

2. Willox, A.c., Stephenson, E., Allen, J., Bourque, F., Drossos, A., Elgaroy, S., Kral, M.J., Mauro, I., Moses, J., Pearce, T. (2015). Examining relationships between climate change and mental health in the Circumpolar North, *Reg. Environ. Chang*, *15*,pp.*151*–163.
3. Altizer, S., Ostfeld, R., Johnson, P.T.J., Kutz, S., Harvell, C.D. (2013). Climate change and infectious diseases: from evidence to a predictive framework. *Science*, 341, pp.514–519.
4. Bouzid, M.,Colon- Gonzalez, F.J., Lung, T., Lake, I.R., Hunter, P.R. (2014). Climate change and the emergence of vector borne diseases in Europe: case study of dengue fever. *BMC Public Health* ,14
5. Chowdhury, F.R., Ibrahim, Q.S.U ., Barl, S.M.D. (2018). The association between temperature, rainfall and humidity with common climate-sensitive infectious diseases in Bangladesh. *PLoS ONE* . 13 (6).
6. Moriyama, M., Hungentobler, W.J., Iwasaki, A. (2020). Seasonality of respiratory viral infections, *Annu. Rev. Virol*.
7. Kudo, E., Song, E., Yockey, L.J. (2019). Low ambient humidity impairs barrier function and innate resistance against influenza infection. *Proc. Natl. Acad. Sci*. *U.S.A,* 116 ,pp.10905–10910.
8. Marr, L.C., Tang, J.W., Van, Mullekom, J., Lakdawala, S.S (2019). Mechanistic insights into the effect of humidity on airborne influenza virus survival transmission and incidence. *J. R. Soc. Interface* ,16.
9. Lowen, A.C., Steel, J. (2014). Roles of humidity and temperature in shaping influenza seasonality. *J. Virol*, 88, pp.7692–7695.
10. Chan, K.H., Peiris, J.S., Lam, S.Y., Poon, L.L., Yuen, K.Y., Seto, W.H. (2011). The effects of temperature and relative humidity on the viability of the SARS coronavirus. *Adv. Virol.*
11. Bukhari, Q., Jameel, Y. (2020) .Will coronavirus pandemic diminish by summer? SSRN [Preprint] Availableat https://papers.ssrn.com/sol3/papers 2020 cfm? abstract_id=3556998.
12. Manogaran, G., Lopez, D. (2017).A Gaussian process based big data processing framework in clustering computing environment, *Springer Science – Business Media*.
13. Ostfeld, R.S., Brunner, J.L. (2015). Climate change and Ixodes tick-borne diseases of humans. *Philos. Trans. R. Soc*, 370.

14. Rodo. X., Pascual ,M., Doblas-Reyes, F.J., Gershunov, A., Stone, D.A., Giorgi, F., Hundson, P.J., Kinter, J., Rodrigue-Arias, M.A., Stenseth, N.C. (2013). Climate change and infectiouos diseases: can we meet the needs for better prediction? .*clim. chang*, 118, pp. 625-640.
15. Tian,H.Y., Zhou, S., Dong, L., Van, T.P., Boeckel, Cui, Y.J., Wu, Y.R., Cazelles, B., Huang, S.Q., Yang, R.F., Grenfell, B.T., Xu, B. (2015). Avian influenza H5N1 viral and bird midration networks in Asia,.*Proc. Natl. Acad. Sci. U. S. A*, 112 , pp.172–177.
16. Kokaze, A., Yoshida, M., Sekine, Y. (2001). The Magnitude of variation in temperature within a year has an effect on the seasonal variations of chickenpox incidence in Japan. *Journal of, Epidemiol. Infect*, pp.269–277.
17. Juhyeon, K., Insung, A. (2021). Infectious disease outbreak prediction using media articles with machine learning models. Scientific reports, Nature.
18. Jabbar M.A., Deekshatulu B.L., Chandra P. (2015). Prediction of heart disease using random forest and feature subset selection. Advances Intelligent Syst Comp Innovations,BioInspired Comp App, pp.187–196.
19. Steyerberg, E. (2009). Clinical Prediction Models: a Practical Approach to Development, Validation, and Updating, Springer Science & Business Media, New York, NY.
20. H. Zou., T. Hastie. (2005). Regularization and variable selection via the elastic net, J. R. Stat Soc. Ser. B, 67 (2), pp. 301–320.
21. Lopez D., Gunasekaran M.,Murugan S.B., Kaur.H. (2014). Spatial Big Data Analytics of Influenza Epidemic in Vellore, India. IEEE International Conference on Big Data.
22. Gössling S, Scott D, Hall C.M. (2020). Pandemics, tourism and global change: a rapid assessment of COVID-19. Journal of Sustainable Tourism.
23. Lopez D., Sekaran G. (2016). Climate change and disease dynamics - A big data perspective. International Journal of Infectious Diseases 45S , pp: 1–477
24. Alfred R., Obit J.H. (2021). The roles of machine learning methods in limiting the spread of deadly diseases: A systematic review. Heliyon, 7(6).