The Research Monograph Series in Computing, Electrical & Communication Networks
Vol. 1 (02), February 2024, pp. 35–40

4 DATA ANALYSIS AND INTERPRETATIONS

Analysis of data pertinent to the work carried out using machine learning and deep learning algorithms for dengue and COVID-19 are presented in this chapter along with the relevant and significant interpretations.

4.1 DATA ANALYSIS

The dengue cases in India were examined using deep learning algorithms. The LSTM network hidden layer as well as the ReLU is applied in this proposed model. The input data contained of latitude and longitude of a geographical space and also the density of population during 2015 to 2019. Other features of data set include area of geography, humidity, temperature and rainfall.

Average rainfall was considered for the period of 2015 to 2019. Every year, the number of dengue cases and impact on human lives in terms of number of deaths were analysed. The data was pre-processed so it is suitable for the application of algorithms. The data collected based on 2011 census and the reference is NIC portal (Khajanchi Subhas and Sarkar Kankan 2006).

The performance metric considered for this study after training of the LSTM-RNN based network is Root mean square value (RMSE). RMSE value for different iterations was computed. The interpretation of the result obtained is RMSE is reducing in hidden layer computation with the increased number of iterations, which is logically apt.

A comparative study of RMSE values with other deep learning algorithms such as gradient boosting, GAM was also performed. It was observed that LSTM's performance is better than other models as the RMSE is of less value (Danane et al., 2021). Therefore, LSTM was considered and the training of the data set related to dengue cases of States and Union territories of India was carried out. The prediction of dengue cases in 2019 across India and the actual dengue cases were analysed and plotted.

The successful training of this proposed model to the huge dataset resulted in 89% of accuracy in number of dengue cases and 82% accuracy, when it comes to deaths due to dengue infected cases. This level of

prediction even it is near around 95% assists the local bodies, state and central governments to be proactive in handling this contagious disease.

Further, the accuracy shall be improved by considering the data set for a longer period of time. If datasets are considered for more than two and half decades, then much better accuracy level attainment is possible. This is another important interpretation from the results obtained.

Other interpretations from the analysis of the graphical representations include; increasing the number of epochs and also the activation functions. Another possibility of arriving at better results lies in executing this model in TPU hardware, whereas the present model was executed on GPU (Singh et al., 2021).

The initial phase of this work was carried out in the Indian context, therefore, the spread of COVID-19 study also considered in the same geographical area. The knowledge and the interpretations drawn in the dengue case study were applied in training of the proposed model to predict COVID-19 spread (Sahoo et al., 2021).

The proposed model is also constrained to the weather and density of population in India. There is another reason to choose this geographical space as India has high value of density of population (Gordon et al., 2020).

4.2 INTERPRETATIONS

The interpretations of the data analysis are also leading to some interesting observations.

(1) India being a diverse country with different climatic conditions, lifestyles of people pertinent to food and healthy habits require a robust study.
(2) The study was conducted for the period of January 2020 and December 2020 (when the pandemic first wave is at onset, peak and settle down), this has given scope to understand the complete spectrum of a wave. The interpretation of the analysed data will be fruitful to be proactive during the second and subsequent waves. The outcomes of the analysis were also shared with local health officials; those helped the health officers in planning vaccination and enhanced testing measures in the Sholinganallur, sub-urban area of City of Chennai.
(3) Month wise infections analysis was helpful in understanding the exact peak time of the Pandemic; incidentally it was during

September of 2020. This data has given a better idea of reaching peak stage after the onset of subsequent waves of the pandemic.

(4) Infected – Deceased analysis w.r.t to Humidity, Temperature and Wind was studied. The data analysis reveals that the virus spread took place in the regions of high temperature and humidity, incidentally states of Tamilnadu and Maharashtra. In fact, these two states are most affected states during first wave of the Pandemic. This outcome also helped in sharing the information with health officials. Interestingly this data is a deviation from the expectation of few researchers across the globe, having an opinion that spread is more in cold climatic conditions rather than hot and humid regions. This is the very nature of Pandemic, always challenging our cognitive skills with its variants.

(5) Weather information, Population density and COVID-19 cases information in India are given as the inputs to the proposed deep learning model. This was helpful in reaching out a comprehensive understanding about various dimensions and their impact in this study. LASSO regression model is used for the relationship study. The data interpretations are more useful for suggesting preventive measures. It was concluded that the threshold for the temperature is set to 80 degrees Fahrenheit and the humidity is set to 70 %, in which case the infected people number is less.

(6) The total study was categorized into Model A and Model B. Model A was related to infected cases and Model B is about deceased cases. This approach has paved the way of crystal clear analysis of the data and in arriving at logical interpretations supported by the comparative studies. The parameters are set for high level of accuracy are learning rate at 0.0005, number of hidden layers at 8, epoch is set to 500, and the time step is at 5. It was observed that proposed model predictions are close to actual numbers in both Model A and Model B investigations.

(7) The performance metrics considered are MSE, RMSE, R-Squared and MAE. These metrics are chosen to find the most accurate deviations in actual and predicted outcomes. The evaluation was carried out for RNN, BRNN, LSTM and Bi-LSTM models. Analysis of the data lead to the interpretation that BRNN model yields better performance than other models.

(8) Another significant interpretation is the results would also have been influenced by the enforcement of lockdown, social distancing, comorbidity issues, release of lockdown and post-lockdown effect

and most importantly early vaccination drives started from the second half of the year 2020.

(9) Tracing, testing and isolating was the method followed to handle in the initial phase of the pandemic in India. But, after the first wave is in the decline, the availability of vaccines Covieshield and Covaxin improved the conditions and reduced the number of cases as well as deceased. But the second wave was a surprise element, which perhaps may be the consequence of the lethargic component of civic society and administrative bodies , ignorant to early warning signals by the prediction models.

The importance of prediction models and mathematical models developed, tested and validated across the globe were so significant during Pandemic. The results and interpretations drawn in this research also give better insights in the data analysis. This work helps to the researchers in exploring more sophisticated and precise modeling of the data. The inclusion of more variety and nature inspired algorithms would produce more accurate results and helps to boost our healthcare and administrative systems.

4.3 CONCLUSION AND FUTURE SCOPE

The research work presented emphasizes the necessity of such studies in these pandemic times with relevance to Indian Sub-Continent. Deep learning model is developed in this work by considering aspect of weather as focal point. There are other set of data related to contagious diseases such as dengue was considered in the first phase of the work, the study carried out with data sets related to 2014-2019 period. The extensive analysis and interpretations of the study was carried out. The work reflected the impact of contagious diseases on humans; this is supported by the study in the case of dengue. Here LSTM, which is a category of RNN, has exhibited better performance over other algorithms.

In the next phase of the work, the experience gained during the initial study was applied in training the data sets related to the on-going Pandemic COVID-19. The study is focussed on cases in India with the real-time data sets belonging the pandemic onset and progressed year of 2020, followed by subsequent training using deep learning model and analysis. The proposed deep learning based model comprised of five networks, namely, Concurrent Neural Network, Recurrent Neural Network, Bidirectional RNN, Long Short-Term Memory and Bidirectional LSTM. The features considered include the weather, recovered patients, infected patients and deceased patients' data. The

outcome is the prediction of raise in number of cases. A comprehensive analysis has been carried out for all possible reasons for the spread of the Pandemic and also prevention, containment measures impact was discussed. A comparative study also done with all the deep learning approaches and found that BRNN algorithm shows better performance as a prediction model. Various performance metrics such RME are considered to understand the relations between iterations in the training and the aberration.

The analysis and inferences also suggesting the reason for LSTM better yield in contagious disease such as dengue and BRNN better performance in the case of COVID-19. LSTM works generally on time series data, the data sets obtained during a period of three years was trained in a better manner naturally by LSTM. Whereas, in the case of COVID-19, the Pandemic is changing variants and creating waves that is enforcing a dynamic state. This is challenging and requires reverse engineering approach most of the times.

The impact has to be studied to predict or the prediction is done based on the already created impact in geography. In this work, BRNN algorithm is used and found better as the causes are evident from the aforementioned reasoning. It can also be studied and tested in future, to explore whether hybrid models leads to more effective prediction and decision making.

4.3.1 Future Scope

This work shall be carried out to understand the dynamic nature of different waves that comes as a part of pandemic. The exact timing and nature of characteristics of different waves will be useful for the medical research and also these predictions will alert the civilians and governments towards proactive measures to be taken during the waves in pandemic.

Further, Intelligent algorithms such as ACO, PSO and ABC shall be used to predict the spread of contagious diseases or a Pandemic such as COVID-19. This will become a potential research prospect as Artificial Swarm Intelligence is an emerging promising topic of interest, which is having extended applications in energy, health and e-commerce sectors. A combination of intelligent algorithms and Deep learning algorithms, in other words, ASI and AI will be a more interesting potential model for predicting the spread of Pandemic. The capacity of training also expected to be more, as features become inclusive and the outcomes become exclusive and interesting.