American Journal of
Electronics & Communication

R Sravan Kumar et. al., American Journal of Electronics & Communication, Vol. II (4), 21-25

Pandemic Analysis: An ANN based Approach

R Sravan Kumar
Department of Computer Science and Engineering,
University of Engineering and Management, Kolkata.
ravishravankumar43@gmail.com

Rajendrani Mukherjee
Department of Computer Science and Engineering,
Universityof Engineering and Management, Kolkata.
rajendrani.mukherjee@uem.edu.in

Abstract - Machine Learning has seen prominent application in healthcare domain. In this paper, an attempt has been made to predict the spread of a disease using machine learning algorithms. In this research, a model is built to predict the casualties created by a particular virus. A pandemic spread analyzer is designed to estimate the cases across the globe. Hyper Parameter Tuned Artificial Neural Network is utilized to implement this research. AI based cognitive modeling simulates properties of biological neural networks. This research implements this model for better preparation to fight pandemic in future.

Keywords – network, pandemic, accuracy, dataset, prediction

1. Introduction

An epidemic is the rapid spread of a disease. It normally happens within a short period of time. During 6^{th} century – 8^{th} centuries, the first plague pandemic jolted the world. The second plague pandemic (14th century–early 19th century) (termed as Black Death) was also precarious and took away many lives. Tuberculosis (TB) became epidemic in Europe in the 18th and 19th century. Thus, epidemic has become a threat in repeated instances. Inthis research, it is explored how machine learning can help in pandemic prediction and possibly help in saving mankind. Early estimation of the pandemic will help in minimizing damages.

Machine learning has seen prevalent usage in building Clinical Decision Support system [8]. ML classifiers help in disease detection and extrapolation. A survey done by Ramalingam et al. in 2018, has investigated how ML classifiers increase the accuracy in predicting diseases [1].Several studies are performed using ANN and deep learning to analyze COVID – 19 pandemic[2].

In this paper, an effort is made to compute the speed at which a virus can spread. The

population size, number of carriers etc. are given as input. Artificial Neural Network is leveraged for this purpose. ANN was proposed by Alexander Bain and William James in the year of 1890 with an idea to imitate the brain. Just like human brain has neurons which are interconnected to each other, ANN has various layers of network which are linked to each other. ANN has seen wide spread application in the field of speech recognition, medical image processing etc. It is expected that the ANN will make decisions in a human-like manner. ANN architecture is built by encoding computers so that they can act like connected brain cells.

2. Related Work

Parthiban and Subramanian leveraged fuzzy logic for intelligent disease diagnosis [3] in 2008. In 2015, Shinde et al. used Naïve Bayes classifier and k-means clustering for categorizing heart disease [4]. In 2016, Li et al. used a combination of KNN and CFS (Correlation Feature Selection) to study EEG graphs[5].

Khanday et al. utilized ensemble classifiers for classifying clinical reports [6]. The authors showed that logistic regression and NaïveBayes performed better by getting 96.2 % testing accuracy. Randhawa et al. tried to analyze COVID – 19 virus genome signatures [7]. The authors united machine learning and MLDSP for categorizing COVID – 19 genomicsequences.

Inspired by all these work, this study explored the application of ANN in virus spread detection.

3. Implementation

The dataset was taken from machine learning repository. Individual data of the pathogens, their rough estimation of sizes and shapes are present in the dataset. The database of epidemic and pandemic lists, their duration and start year were created. The model starts with datacleaning and filling in the missing values with mean values.

The process of binning is used to create bins for the parameter of Death Toll. Different categories of this parameter are –

- Outburst Class 1(Cases=0-1K),
- Outburst Class 2(Cases=0-1K),
- Epidemic (Cases=10k-100k)
- Pandemic(100k+) etc.

Fig.1 represents the dataset in detail. Important fields from the dataset are – disease name, disease type, carrier etc.

Next, the ANN was implemented. In case of ANN, the input layer accepts inputs in several formats provided by the programmer. The hidden layer performs all the calculations tofind hidden features and patterns. The input has gone through transformations with the help of hidden layer and finally output result is obtained. The weighted sum of the inputs and a bias is represented as –

$$\sum_{i=1}^{n} W_i * Y_i + \mathrm{B} \qquad (1)$$

The proposed model will be taking the given inputs and will classify the virus on the probabilistic value on how much cases can be caused due to the spread of virus in the given categories. The model is created using Artificial Neural Networks by forming layers which have different no. of neurons present in it so as to predict the output in the output layer with the probabilistic values generated using the weights and parameters fed to the network (Fig.2). Bayesian Optimizer is used which helps in setting the most appropriate parameters for the model.

DiseaseName	TimeDurationInYears	Areas	DiseaseType	SubType	DeathToll	Carriers	MinSize	MaxSize	Medium	StartYear
Babylon Influenza	3	Continent	Virus	Influenza		2	0.08	0.12	Air	-1200
Plague of Athens	3	Country	Bacteria	Typhi	100000	1	1	3	Water	-429
412 BC Epidemic	1	Continent	Virus	Influenza		2	0.08	0.12	Air	-412
Antonine Plague	35	Continent	Virus	Variola	10000000	3	0.244	0.27	Air&Water	165
Jian'an Plague	1	Country	Bacteria	Typhi		1	1	3	Water	217
Plague Of Cyprian	16	Continent	Virus	Variola		3	0.244	0.27	Air&Water	250
Plague of Justinian	8	Continent	Bacteria	Bubonic	100000000	5	2	20	Direct	541
Roman Plague	1	Country	Bacteria	Bubonic		5	2	20	Direct	590
Plague Of Sherqe	2	Country	Bacteria	Bubonic	25000	5	2	20	Direct	627
Plague of Amwas	2	Worldwide	Bacteria	Bubonic	50000	4	2	20	Direct	638
Plague of 664	25	Country	Bacteria	Bubonic		5	2	20	Direct	664
Plague of 698	4	Continent	Bacteria	Bubonic		4	2	20	Direct	698
Japanese Smallpox	3	Country	Virus	Variola	200000	3	0.244	0.27	Air&Water	735
Plague of 746	2	Continent	Bacteria	Bubonic		4	2	2.5	Direct	746
Black Death	8	Worldwide	Bacteria	Bubonic	20000000	4	2	20	Direct	1346
Sweating Sickness	66	Continent	Virus	Hanta	10000	2	0.12	0.16	Water	1485
1489 Spain Typhus	1	Country	Bacteria	Typhi	17000	1	1	3	Water	1489
1510 Influenza Pandemeic	2	Worldwide	Virus	Influenza	1000000	2	0.08	0.12	Air	1510
1520 Mexico smallpox	3	Country	Virus	Variola	8000000	3	0.244	0.12	Air&Water	1519
Cocoliztli Epidemic	4	Country	Bacteria	Enterica	15000000	5	0.02	0.05	Direct	1545
1557 Influenza pandemic	3	Worldwide	Virus	Influenza		2	0.08	0.12	Air	1557
1561 Chile smallpox epidemic	2	Country	Virus	Variola	500000	3	0.244	0.27	Air&Water	1561
1563 London plague	2	City	Bacteria	Bubonic	21000	4	2	20	Direct	1563
Cocoliztli epidemic of 1576	5	Country	Virus	Enterica	2500000	5	0.02	0.05	Direct	1576
1582 Tenerife plague epidemic	2	City	Bacteria	Bubonic	9000	4	2	20	Direct	1582
1592–1596 Seneca nation measles epidemic	5	Continent	Virus	Morbillivirus		2	15	16	Air,Water& Direct	1592
1592–93 Malta plague epidemic (part of the Second plague pandemic)	2	Country	Bacteria	Bubonic	3000	4	2	20	Direct	1592
1592–93 London plague (part of the Second plague pandemic)	2	City	Bacteria	Bubonic	20000	4	2	20	Direct	1592
1596–1602 Spain plague epidemic (part of the Second plague pandemic)	7	Country	Bacteria	Bubonic	700000	4	2	20	Direct	1596
1600–1650 South America malaria epidemic	50	Continent	Parasite	Malaria		2	10	15	Direct	1600
1603 London plague epidemic (part of the Second plague pandemic)	1	City	Bacteria	Bubonic	40000	4	2	20	Direct	1603

Fig. 1: Different Fields of Subject Dataset

```
In [38]: params_nn_ = nn_bo.max['params']

         learning_rate = params_nn_['learning_rate']
         activationL = ['relu', 'sigmoid', 'softplus', 'softsign', 'tanh', 'selu',
                        'elu', 'exponential', LeakyReLU,'relu']
         params_nn_['activation'] = activationL[round(params_nn_['activation'])]

         params_nn_['batch_size'] = round(params_nn_['batch_size'])
         params_nn_['epochs'] = round(params_nn_['epochs'])
         params_nn_['layers1'] = round(params_nn_['layers1'])
         params_nn_['layers2'] = round(params_nn_['layers2'])
         params_nn_['neurons'] = round(params_nn_['neurons'])

         optimizerL = ['Adam', 'SGD', 'RMSprop', 'Adadelta', 'Adagrad', 'Adamax', 'Nadam', 'Ftrl','Adam']
         optimizerD= {'Adam':Adam(lr=learning_rate), 'SGD':SGD(lr=learning_rate),
                      'RMSprop':RMSprop(lr=learning_rate), 'Adadelta':Adadelta(lr=learning_rate),
                      'Adagrad':Adagrad(lr=learning_rate), 'Adamax':Adamax(lr=learning_rate),
                      'Nadam':Nadam(lr=learning_rate), 'Ftrl':Ftrl(lr=learning_rate)}
         params_nn_['optimizer'] = optimizerD[optimizerL[round(params_nn_['optimizer'])]]

         params_nn_

Out[38]: {'activation': 'softsign',
          'batch_size': 246,
          'dropout': 0.41199620026358696,
          'dropout_rate': 0.21157789620747203,
          'epochs': 120,
          'layers1': 9,
          'layers2': 4,
          'learning_rate': 0.07804182844303093,
          'neurons': 81,
          'normalization': 0.8148417763675094,
          'optimizer': <tensorflow.python.keras.optimizer_v2.nadam.Nadam at 0x7f8b841f3be0>}
```

Fig. 2: Code Snippet for ANN Implementation

4. Results

The obtained results are tabulated in Table 1.

The Precision is represented by the ratio of True Positive elements to total number of positively predicted units. True Positive are the elements that have been labeled as positive by the model and they are actually positive, while False Positive are the elements that have been labeled as positive, but they are actually negative.

Precision is a good measure to determine, when False Positive is high. For instance, in email spam detection, user may lose track of important emails if the precision is not high for the spam detection.

Recall determines how many of the True Positives the model can detect. F1 score is the harmonic mean of precision and recall. In order to achieve a balanced classification, we have tried to maximize the F1 score.

Table 1: Obtained Results after ANN run

Metrics	Value
Accuracy	81%
F1-Score	92%
Precision	63%
Recall	62%

5. Conclusion

The research is conducted keeping in mind the current pandemic scenario. According to the October 2020 'Era of Pandemics' report by the UnitedNations' Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (written by 22 experts in a variety of fields) it is estimated that as many as 850,000viruses will be transmitted from

animals to humans. Hence, towards an attempt to analyze the propagation and prevention of pandemic this research utilizes the cutting edge technologies.

References

1. Ramalingam, V. V., Dandapath, A., &Karthik Raja, M. (2018). Heart disease prediction using machine learning techniques: A survey. *International Journal of Engineering and Technology(UAE)*, *7*(2.8 Special Issue 8), 684–687. https://doi.org/10.14419/ijet.v7i2.8.10557
2. Amar, L. A., Taha, A. A., & Mohamed, M. Y. (2020). Prediction of the final size for COVID-19 epidemic using machine learning: A case study of Egypt. *Infectious Disease Modelling*, *5*, 622–634. https://doi.org/10.1016/j.idm.2020.08.008
3. Parthiban, L., & Subramanian, R. (2008). Intelligent Heart Disease Predic- tion System using CANFIS and Genetic Algorithm. *International Journal of Biological and Medical Sciences*, *3*(3), 157–160.
4. Shinde, R., Arjun, S., Patil, P., &Waghmare, P. J. (2015). An Intelligent Heart Disease Prediction System Using K-Means Clustering and Naïve Bayes Algorithm. *International Journal of Computer Science and Information Technologies*, *6*(1), 637–639.
5. Li, X., Hu, B., Sun, S., &Cai, H. (2016). EEG-based mild depressive detection using feature selection methods and classifiers. *Computer Methods and Programs in Biomedicine*, *136*, 151–161. https://doi.org/10.1016/j.cmpb.2016.08.010
6. Khanday, A. M. U. D., Rabani, S. T., Khan, Q. R., Rouf, N., &MohiUd Din, M. (2020). Machine learning based approaches for detecting COVID-19 using clinical text data. *International Journal of Information Technology (Singapore)*, *12*(3), 731–739. https://doi.org/10.1007/s41870-020-00495-9
7. Randhawa, G. S., Soltysiak, M. P. M., El Roz, H., de Souza, C. P. E., Hill, K. A., & Kari, L. (2020). Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. *PLoS ONE*,*15*(4), 1–24. https://doi.org/10.1371/journal.pone.0232391
8. Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2013). Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm. *Procedia Technology*, *10*, 85–94. https://doi.org/10.1016/j.protcy.2013.12.340