International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 08 - Issue 03 || March 2022 || PP. 01-05

Artificial Neural Network and Big Data approaches to predict oxygen demand to combat Covid-19

Nitiraj singh Sandu[1], Shakir Karim[2]
[1]*AIBI Sydney raj.sandu@aibihe.edu.au*

Abstract: COVID-19 outbreak is the biggest challenge in today's world. It became a nightmare of this century not only for healthcare, specialist but also for governments, business circles, and many other professions. So far millions of humans have lost their life by being victim of this disease. To combat this pandemic, along with other sectors, computer scientist also came into this field. Leveraging progress of artificial intelligence, computer scientist played a great part to fight against COVID-19. In this research we will present novel framework for investigating oxygen demand for COVID patient. In this technology framework we will use Artificial Neural Network (ANN) and big data to predict the oxygen demand of COVID patient. Result of this framework will help government sector to make decision regarding oxygen demand and supply. Since India is country in the world currently that is suffering from oxygen unavailability, resulting in many deaths. We had chosen India as our case study, but our framework is generic that can be implemented in any partof the world. We believe our proposed framework will play a great part in combating this fatal outbreak.
Keywords: COVID 19., Artificial neural network(ANN), Big data, Oximeter.

I. Introduction

History of mankind is filled with immense number of disasters. Most of these disasters are not man mad rather they belong to category of natural disasters, not only an accident is the cause of such kind of tragedy but also pandemics also disturbed mankind life in many ways. [1]. In the last decade mankind witnessed the worst emergence is a virus, that spreads across the borders and crosses, oceans, and continents, ultimately wrap up all over world, is named Corona Virus Dieses 2019 (COVID19). COVID-19 start from city of Wuhan, China in late 2019, s and in current time it has been spread over 250 countries of the world. Millions of precious lives have been lost from this disease. Healthcare emergency has been set in almost all countries in the world. Although numerous vaccines have come in the field yet there is still no final cure of this disease has discovered. All governments and private health care organizations are struggling their best to battle against this curse. Since it is a respiratory disease, there comes a stage of this illness, where patient need oxygen to save their life. Due to huge number of patients, oxygen availability strictly needs to be made available. Recently some countries are suffering from short fall of oxygen, needed for COVID patients. In our research we will play our part to investigate oxygen demand by using cutting-edge technologies. oximeter is a device that is used to measure the ratio of oxygen in blood. Pulse oximeter has proven great beneficiary in detecting oxygen demand in an individual body.

Data from Pulse oximeter can be sent to server through internet. To overcome this huge amount of data coming from millions of human's bodies is difficult to handle from simple relational database management system. Big data come into field to handle needs of managing huge data. Big data is used to manage massive amount of data, since in our search huge amount of data needs to be handled, we will leverage big data features to store, process and output our data. Big data analytic as its name refer is field of study and analyze big data. After collecting humans body data, there comes the stage of processing this data to make useful decision. Thanks to Artificial Intelligence (AI), that gives us techniques of such a complex processing. There comes a narrow field AI that called machine learning. Machine learning algorithms are designed in such a way that they can train their self from the input data [2]. Machine learning have following categories. In supervised learning data is given and our goal is to apply find output from any algorithm. Most common type of supervised learning algorithms classes are classification and regression. In contract to supervised learning in unsupervised learning no clean dataset are available rather it predicts result based to pre-defined pattern. Reinforcement learning is different from both of above, as in reinforcement learning algorithm is designed in such a way that it learns from itself by its environment. In our research we will use Artificial neural network algorithm of machine learning to detect oxygen demand in current situation. This decision will send to government sector agencies so that they can take proactive measure to avoid shortfall of oxygen demand before time. If oxygen demand makes sure it can save many precious human lives.

Since India is most effected country in the world that is suffering from oxygen shortfall, so we take India as use case. Social media and other news sources show that in India thousands of people are dying due to unavailability of oxygen.In remaining part of our paper, first we present In remaining part of our paper, first we

resent related work of technologies used in our framework, section III will present research methodology, section IV will describe research experiment and at the last we will have section of conclusion along with research direction for future work.

II. Related work

Biosurveillances a branch of healthcare that deals with detection and countermeasure of threats to human life. In this area of research harmful diseases, particularly, epidemics and pandemics are detected, and try to diminish threats to human life from such diseases [3]. Along with other bio informatics technologies, machine learning, big data analytics and natural language processes are the most common technologies uses in bio surveillances [4].To detect and controlling pantomimic outbreaks, they analyze, social media news, health care reports analysis and other current updates for draw attention to any pandemics before it outbreaks[5]. Blue Dot, A Canadian company, first detect COVID pandemics in early December 2019.[6, 7].Data from sources such as medical reports, satellite images, (rush near hospitals) and other sources are analyzed through big data analytics[8, 9].“Big Data” deals with massive amount of data sets, that are hard to process and analysis with traditional database management systems. The major properties of big data are usually referring to as three V, s, that is Velocity, Volume and Versatility. The term volume refers to huge amount of data to be stored and handled. Velocity, as its name suggest refers to speed of its input and output like how speedily data can be stored and retrieved. Variety is its feature of dealing with different nature of data sources. And finally, veracity is its ability of accuracy of data processing. “Big Data Analytics” is a name of technique in which big data is processed in such a way that it can be helpful to gather useful information [10].

Machine learning is most popular area of artificial intelligence in today’s word, Machine Learning is a branch of AI in which data set is trained in such a way that it can learn from its own experience. Neural Networks are branch of machine learning that deals with algorithms that work similarly as human neurons. Artificial Neural Network (ANN) are special kind of neural networks that have widely be used in epidemiological marvels.ANN works like human brain. Moreover, ANN is also used to predict pandemic peaks risk from this pandemic can be analyzed [11].Gohil[12] proposed a model, that applied ANN to predict number of recovered cases and death occurrence from that disease. Oximeter is widely adopted device in recent years, its purpose is to detect oxygen saturation level in human body [14-15]. This machine finds oxygen saturation level (SPO2) with zero harm level to human body. It can detect SPO2 level and can trigger in case of decreasing this level from threshold level. Wirz[16] presented a framework for detecting individuals SPO2 level to find that either they are suffering from COVID or not. In our best knowledge there is no single work that use Oximeter data to analyze oxygen demand. Also, ANN is used in forecasting patience number etc. as mentioned above but it is not being used with big data analytic to find oxygen demand in a city or a country. Hence, we proved that our work is novel, and no such work has been done earlier. In our next section we will provide our proposed model.

III. Research methodology

This section include framework that we proposed for detecting oxygen demand by using big data and ANN. We have phases of framework that is elaborated in next part of this section.

A. Data Input

In our framework pulse oximeter will used as front-end device that will be attached in human body. To detect SPO2 ratio of an individual we will need following equipment’s

- MAX30100_PulseOximeter
- ESP8266WiFi.h
- JHD 162A 16*2 LCD display
- Arduino microprocessor
- Connection wires

Arduino is one of the most famous microprocessors used by electronic professionals. With help of Arduino, we can develop a program very easily, with help of its available libraries, without much prior knowledge of embedded system programming [19]. Arduino IDE is used for coding in C langue. This code will send reading of oximeter to our server using API. Fig 1 is describing prototype of pulse oximeter configuration with Arduino.

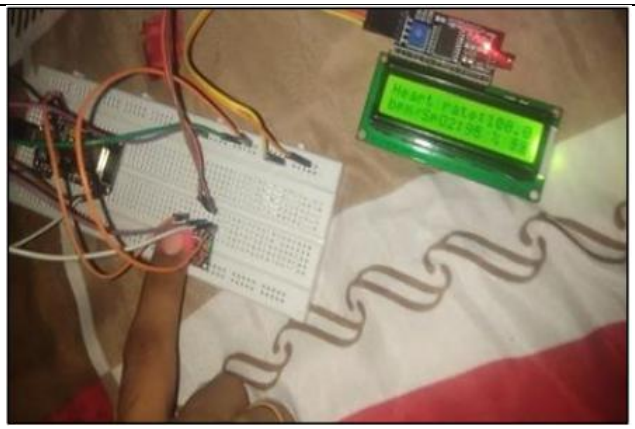

Fig 2: prototype of pulse oximeter [16]

B. Data Normalization

In this step values taken from earlier step will change to a common scale. This is necessary to make our dataset in such a form that we can process and apply algorithm on it.

We will set values on linear scale rather than values of values on a linear scale level rather than different ranges. To make data normalize we will use term of min-max. Following formula is being used to scaling the data $Y\ scaled = y - \min(y)\ \max(y) - \min(y)$ [17].

C. Data alignment/aggregation:

Data comes from different sources, that in our case are oximeters, are needed to merge in a single dataset. This step will lead us to final data set on which our algorithm work. We take care of timestamp of data to be stored.

D. Data formulation:

At this stage big data role starts to play its part. Since, our proposed mechanism will store substantial amount of data, it will need to formulate from big data tools and techniques. We will use MongoDB as big data management system tool. We will store the data come from user's body sensors in form of collections in MongoDB. These collections will
make a dataset from synthetic data generation in next stage.

E. Synthetic data generation

Data generated from above mentioned steps are now need to pretrain. For this purpose,
Synthetic data needs to be generated. To generate synthetic data set we used TUTOR framework [18]. TUTOR framework generates synthetic data set by following three ways.

- Multi-variant normal distribution: In this approach data comes from real data sources, in our case data comes from oximeters, are considered as normal
- distributed for synthetic data.
- Gaussian Mixture model: In this model validation data set is used to obtained optimal GMM.
- Kernel Density Estimation: In this approach synthetic data is generate by Calculating sum of kernels.

E. Building a Knowledge Base

In this step we will label data point. To build knowledge base we will use commonly used machine learning algorithms. We choose Random Forest (RF) and Decision Tree (DT) as algorithm to be used for generating knowledge base.

F. Training synthetic data

In this step we will train synthetic data. To achieve this, we pre-train the ANN model by using synthetic data generated in earlier Steps. Figure 2 is describing our proposed framework.

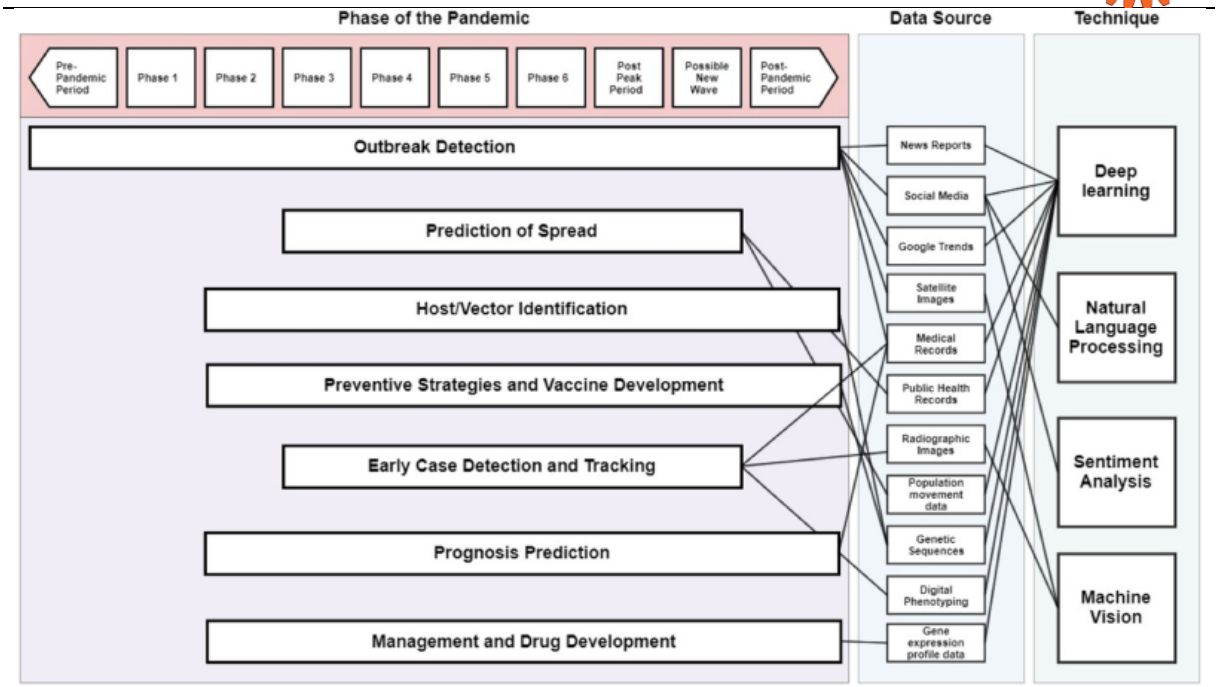

Fig.2. Our proposed framework

In our next section we will present result of our experiment.

IV. Experimental Result

This section includes test bed to evaluate performance of after mentioned framework. In this analysis we make three ways of classification for the coherent describe earlier. Furthermore, we accomplish a study to analyses different use cases of detecting oxygen demands. To evaluate our model, we designed following different kind of metrics

- True positive
- False negative rate
- False positive rate

Under mentioned is detail of these metrices:

- True positive (negative): This term will use to check accuracy of alert generated for oxygen short come.
- False positive (negative): These metrics will used to determine ratio on which our experiment generate exact.

These matrices will help to evaluate performance of our framework from different perspective. Test accuracy will determine overall accuracy of our framework. In general, it will be a ratio of accurate result of our experiment to the overall instances. For instance, in case we perform evaluation on 100 person, and 92 results are accurate, like they are suffering from oxygen shortfall, this shows 92 % test accuracy of our framework. False Positive ratio is defined as the ratio in wrong decisions i.e., person that is health and do not need oxygen to the number of right decisions. Same in case of false negative ratio will used to check on Which ratio our experiment provides wrong results. To compare performance of our model We performed ablation study.

V. Conclusion and future work

Outbreak of COVID-19 is a nightmare for humanity from last two years. A massive work has been done in detecting occurrence of this virus in human body. Computer science played a vital role for healthcare department with its cutting-edge technologies such as Artificial Intelligence, Big data etc. In our research we point out a novel aspect in this research that is to determine oxygen demand for COVID-19 patients. Numerous countries, particularly southeast Asian country India, is suffering oxygen short fall. In that research we proposed a framework to determine oxygen demand to help government agencies for making proactive measure for fulfilling oxygen demand. In this research we used data from human bodies with help of oximeter to assess that they need oxygen for healthy life. We use hybrid technologies like oximeter, big data analytics, and artificial neural network to sort out that Phenomena.

We believe that implementation of our proposed framework will facilitate Government to fulfill their oxygen demand. In future we are aiming to perform physical test of our proposed model so that it became near to real life.

VI. References

[1]. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. *Nature*.2008; 451(7181):990–3.

[2]. Wiens J, Shenoy ES. Machine Learning for Healthcare: On the vergeofamajor shift in healthcare epidemiology. *Clin. Infect Dis. Off Publ. Infect Dis. Soc. Am.* 2018 Jan1; 66(1):149–53.

[3]. AllenKC.Chapter13-Applications: Biosurveillance, biodefense, and biotechnology. In: Horney JA, editor. Disaster epidemiology [Internet]. AcademicPress; 2018 [cited 2020 Mar 29]. p.143–51. Available from: http://www.sciencedirect.com/science/article/pii/ B9780128093184000198

[4]. Jalal A. Analytics, Machine Learning &NLP – use in Bio Surveillance and Public Health practice. Online J Public Health Inform [Internet]. 2015 May 7 [cited 2020 Mar 29]; 7(1).

[5]. Al-Garadi MA, Khan MS, Varathan KD, Mujtaba G,Al-Kabsi AM. Using online social networks to track a pandemic: A systematic review

[6]. Bragazzi, N. L., Dai, H., Damiani, G., Behzadifar, M., Martini, M., & Wu, J. (2020). How big data and artificial intelligence can help better manage the COVID-19 pandemic. *International journal of environmental research and public health*, *17*(9), 3176.

[7]. Mc Call B. COVID-19 and artificial intelligence: protecting health- care workers and curbing the spread. *Lancet Digit Health.* 2020 Apr1; 2(4):e166–7.

[8]. Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MUG, Khan K. Pneumonia of unknown aetiology in Wuhan, China: potential for international spread via commercial air travel. *J. Travel. Med.* [Internet]. 2020 Mar 13 [cited 2020 Apr 26]; 27(2). Available from: https://academic.oup.com/jtm/article/27/2/taaa008/5704418

[9]. Funk S, Camacho A, Kucharski AJ, Eggo RM, Edmunds WJ. Real- time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. *Epidemics.* 2018 Mar; 22:56–61.

[10]. Johansson MA, Apfeldorf KM, Dobson S, Devita J, BuczakAL, Baugher B,etal. Anopen challenge to advance probabilistic for e-casting for dengue epidemics. *Proc. Natl. Acad. Sci.* 2019 Nov 26; 116(48): 24268–74.

[11]. Teng Y, Bi D, Xie G, JinY, Huang Y, Lin B, et al. Dynamic Forecasting of Zika Epidemics Using Google Trends. *PLoS One*2017 Jan6; 12(1): e0165085.

[12]. GohilS, Vuik S, Darzi A. Sentiment analysis of healthcare tweets: review of the methods used. *JMIR Public Health Surveill* [Internet]. 2018 Apr 23 [cited 2020 Mar 29]; 4(2). Available from: https:// www.ncbi.nlm.nih.gov/pmc/articles/PMC5938573/

[13]. LimS, Tucker CS, Kumara S. Anun supervised machine learning model ford is covering latent infectious diseases using social media data.

[14]. Choi S, Lee J, Kang M-G, Min H, Chang Y-S, Yoon S. Large-scale machine learning of media outlets for understanding public reactions to nation-wide viral infection out breaks. *Methods San Diego Calif.* 201701; 129: 50–9.

[15]. Wirz CD, Xenos MA, Brossard D, Scheufele D, Chung JH, Massarani L. Rethinking social amplification of risk: Social media and Zikain three languages. *Risk Anal. Off Publ. Soc. Risk. Anal.* 2018; 38(12): 2599–624.

[16]. Gui X, Wang Y, Kou Y, Reynolds TL, Chen Y, Mei Q, et al. Understanding the Patterns of Health Information Dissemination on Social Media during the Zika Outbreak. *AMIA Annu. Symp. Proc. AMIA Symp.* 2017; 2017: 820–9.

[17]. Venna SR, Tavanaei A, Gottumukkala RN, Raghavan VV, Maida AS, Nichols S. A Novel Data-Driven Model for Real-Time Influenza Forecasting. *IEEE Access.*2019; 7:7691–701.

[18]. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modeling study. *Lancet Infect. Dis.* [Internet]. 2020 Mar 11 [cited 2020 Mar 28]; 0(0). Available from: https://www. helancet.com/journals/laninf/article/PIIS1473-3099(20)30144-4/ abstract.