

https://dx.doi.org/10.12785/ijcds/120101

Deep Learning based Analysis of Covid-19 Mortality Risk

Touseef Iqbal¹ and Shaima Qureshi²

¹Department of Computer Science and Engineering, National Institute of Technology Srinagar, J&K, India ²Department of Computer Science and Engineering, National Institute of Technology Srinagar, J&K, India

Received 4 Mar. 2021, Revised 6 Apr. 2022, Accepted 15 Jun. 2022, Published 1 Jul. 2022

Abstract: Viral infectious diseases such as Covid-19 present a major threat to public health. Despite extreme research efforts, how, when and where such new outbreaks appear is still a source of substantial uncertainty. Deep learning (DL) is playing an increasingly important role in our lives. This paper presents one of the popular deep learning technique, Long Short Term Memory (LSTM) for prediction of Corona-Virus cases. The handcrafted feature extraction of traditional methods is less scalable on large data-sets, but deep learning algorithms perform extremely well on large data-sets, because of automatic feature extraction. Deep learning has already made a huge impact in areas, such as cancer diagnosis, precision medicine, self-driving cars, predictive forecasting, and speech recognition. This paper highlights the approaches where deep learning can be helpful to tackle the Covid-19 virus and similar outbreaks. This paper also discusses the structure and functioning of Covid-19. The utilization of different deep learning concepts like Convolutional Neural Networks, Transfer Learning for this pandemic is also highlighted.

Keywords: Covid-19, Deep Learning (DL), Convolutional Neural Networks (CNNs), Long Short Term Memory (LSTM).

1. INTRODUCTION

Corona-Virus is turning out to be the worst viral disease outbreak of all times. It is an irresistible infection brought by newly found Corona-Virus. The Corona-Virus was first identified in a bird in 1937 [1]. The name is derived from its structure which looks like crown spikes along the surface, studied under microscope. Slowly, the study revealed that there is a full family of Corona-Viruses that can infect bats, rats as well as humans. The first human Corona-Virus was discovered in the nose of a person suffering from a common cold in the 1960s. The most recent Corona-Virus called Corona Virus disease 2019 (Covid-19) was originated from seafood market in Wuhan, China in Dec 2019. This Virus has already killed more than 54 lakh people and more than 2.5 million cases are positive as of now all around the world. The severe situation puts forward new requirements for the prevention and control strategy [2]. Many technologies have started to study this pandemic and come up with solutions to slow down this outbreak. For instance, BlueDot, a global database company for Artificial Intelligence (AI), uses AI driven Machine Learning (ML) and Natural Language Processing (NLP) to analyze data from a variety of sources and track over 100 infectious diseases. The sub-field of AI has shown promising results for Deep Learning (DL) in different domain like NLP, Computer Vision (CV), Medical Imaging (MI), HealthCare etc. Deep learning models, such as Convolutional Neural Networks (CNNs), can model nonlinear problems, possess variety of applications, including medical imaging. Despite the fact that this area is not currently researched well, but the ability to determine the unknown structure of input data using low level features to determine high level feature vectors have already solved different problems in many domains e.g. autonomous vehicles, text generation, computer vision, medical imaging etc. However, due to the limitation of training data and inability to incorporate complex structures into deep learning models makes it difficult to use in practice. The improvement in area of unsupervised learning (Generative Modelling) can be very helpful to tackle such issues. The applications of deep learning in medical imaging like MRI, Radiography, Ultrasound shows the diversity of its use in medical field [3]. This paper will highlight the approaches where deep learning can be useful to tackle the problem of Covid-19 and outbreaks like it. In section II we discuss structure and functioning of Covid-19. This is followed by deep learning for diseases diagnosis in section III and section IV discusses the deep learning for Covid-19 diagnosis where LSTM is trained to Covid-19 prediction. Finally conclusion is made in section V.

2. STRUCTURE AND FUNCTIONING OF COVID-19

Corona-Viruses are the large family of enveloped RNA (Ribonucleic acid) Viruses. These Viruses are broadly divided in two Groups:-

- 1) **Alpha and Beta:** these are such types originated from Bats and Rats
- 2) Gamma and Delta: originated from Alien species,



e,g Avian and Pig.

Corona-Viruses are responsible for variety of diseases in many species. They were suspected to cause mild selfrestricting respiratory infections in humans until 2002, when Beta-Corona-Virus crossed species barriers from bats to himalayan palm civet and then humans. This resulted in Severe Acute Respiratory Syndrome (SARS-Epidemic). The Beta-Virus was recorded from 2002-2003 with Fatality Rate (FR) of 9%. FR is the ratio of number of deaths due to some disease to the number of people positive with that disease in given time. Another Corona-Virus called Middle East Respiratory Syndrome (MERS) emerged in 2012. This virus originated from Bats to dromedary camel and then to humans with FR of 40-50%. The Novel Corona Virus 2019 (Covid-19) pandemic is also a Beta Corona-Virus. The virus genome is fully sequenced and appears to be very similar to the strain in bats, meaning that it also originated from bats as well. Therefore, the virus is identical to the SARS Corona-Virus, which is called SARS- Cov-2. It is not clear to date, whether it appears directly from bats to humans or there is an intermediary host. The genome of Corona-Virus is a single stranded positive sense RNA molecule that contains all information necessary to make a viral component. The Nucleocapsid is enclosed in an envelope which is basically a lipid membrane embedded with proteins as shown in figure 1.



Figure 1. Structure of Corona-Virus

The envelope is surrounded with spike like structure which gives the appearance of a crown. The crown in Latin is known as corona, this is where the corona name was originated. For viral infection, the integrity of envelope is necessary and can be damaged by lipid solvents such as soap, alcohol, disinfectants, etc. In-fact envelope viruses, when they are outside the host, are easy to inactivate. In order to infect the host, the virus spike must bind to a molecule called the receptor on the cell surface. The nature of this binding clarifies, why viruses are usually specie specific. There are receptors in certain species but not others [4]. The host jumping is usually triggered by mutation in spike proteins which change them in a way that they can bind the new species. The novel Corona-Virus appears to invade human cells using the same receptor as SARS-CoV-2 and that enzyme is called Angiotensin Converting Enzyme 2. (ACE- 2).

Infection usually starts from cells in respiratory system and then replicates its viral RNA into proteins. These are then assembled in new viral particles called Verions. The host cell dies as the new Verions get released. The uncontrolled growth of virus destroys respiratory tissues. Infection triggers the bodies' inflammatory response which brings the immune cells to fight the virus. Respiratory tissues are damaged by uncontrolled virus growth. Infection activates the body's inflammatory response that enables the immune cells to fight the virus. That is why healthy persons having strong immune system show less infection [2]. The symptoms shown by persons infected with Covid-19 are dry cough, fever and tiredness. Sore throat, runny nose or nasal congestion are some symptoms that many patients might develop. The symptoms are mild and begin slowly. Many infected people neither feel unwell nor develop any symptoms. Majority of infected people (about 80%) recover from this infection without special treatment. The ratio of people who get infected with Covid-19 and become seriously ill is 1: 6. People with pre-existing medical problems like diabetes, heart problems or high blood pressure are more like to develop serious illness. The next section will discuss how deep learning models have been used to minimize the effect of other diseases.

3. DEEP LEARNING FOR DIAGNOSIS OF VIRUSES:

Medical imaging is one of the most important source of diagnostic information, however it depends on human interpretation and subject to increasing resource challenges. Automated diagnosis of these images through deep learning models might provide the solution to address this problem. Deep learning advancements have replaced the complex, human developed agents. Deep learning models have contributed in different ways to slow down the preexisting similar outbreaks e.g. SARS, MERS [5] etc. The most important step of analyzing the virus is to identify its viral sequence from metagenoic samples. Traditional methods for diagnosing the viral sequence uses NCBI BLAST " tool to find out the similarity between sequences". The sequences are compared to the known genomes from public database to calculate the similarity among them. The divergence in these samples makes the algorithm classify most of the sequences as "unknown". Deep learning method as a different approach uses CNNs, which are trained to classify these genomes and generalizes well on new samples [6] [7]. Many deep learning based methods to find out viral sequences were proposed like RSCU [8] (Relative synonyms codon usage frequency), prediction of lytic virus [9] etc. One amoung the important methods for the surveillance of viral disease is to determine its host

[10]. However, the mechanism which is responsible for its adaptation to new hosts is largely unknown. The ability of deep learning models to extract features by themselves, undoubtedly solves many specific tasks [11]. Predicting the potential host of viral disease will be helpful to gain some insights about its history, thus provides directions to slow down its transmission. Deep learning models are applied to predict the potential viral hosts by analyzing the viral genome of diseases. The molecule frequencies used as the feature set, deep learning model (CNN) were able to predict the original hosts of the virus [12]. The factors that determine the influence of virus is one of the approaches used to identify the severity of diseases. Many classification models are able to predict these factors. During MERS-Cov2 viral infection, factors that were determined are age, pre-existing diseases and symptoms [13]. Following the same methodology deep learning models can be very handy to determine the factors which may slow down the Covid-19 pandemic.

A. Early Prediction

Deep learning can help us in many ways to slow down the outbreaks like Covid-19. One among them is having the early prediction of how outbreak can spread and calculation of how curve flattens with the number of confirmed, recovered and death cases by analyzing previous data [14]. Deep learning can be helpful to predict when outbreaks like Covid-19 may occur by analyzing the data of previously recorded outbreaks. In addition to this by analyzing the data, we can also predict how quickly the virus can spread e,g the number of people that might get infected. There are variety of algorithms in deep learning such as Multi-Layer Perceptron (MLP), Recurrent Neural Networks Long Short Term Memory etc. that can be used to model such problems. This not only helps us to predict the fatality rate of the virus in any community, but also makes people aware of taking necessary precautions before such outbreaks might hit the door. Deep learning has the ability to analyze large amount of data and extract insights which can lead us to the deeper knowledge of such pandemic [15]. This also will be useful to enable health and Govt officials to make better preparation throughout the outbreak. Deep learning may not only warn us about looming pandemics, but can also provide insights to identify, develop and extend the treatments and vaccines faster than before. One similar work was proposed by [16], where the number of cases were predicted based on the infected rate of that area in certain period of time.

Recurrent Neural Networks (RNNs) a popular deep learning architecture mostly applied for modeling time series data. The structure of RNNs is different from traditional machine learning algorithms as recurrent nature helps the model to make predictions based on previous output. For example the language model predicting the next word based on previous words. There are few shortcomings in vanilla RNNs (Vanishing and Exploding gradient), so over the years few modifications in the architecture of RNNs were made to mitigate these issues. Long Short Term Memory Networks



(LSTM) an extension of RNNs which uses different internal computations to address these problems. LSTMs use different methods to compute the hidden state. Apart from hidden state each LSTM cell contains a cell state vector. The gating mechanism maintains the cell state in a way, that LSTM vector can read, write or reset from it at each time step. Each LSTM cell contains three gates (same as binary gates) of same shape:-

1 Input Gate: The input gate control the memory cell to update the cell state and the equation is

$$i^{t} = \sigma(w^{i}(h^{t-1}, x^{t}) + b^{i}) \tag{1}$$

2 Forget Gate: The forget gate controls the memory cell to reset the cell state and the equation is

$$f^{t} = \sigma(w^{f}(h^{t-1}, x^{t}) + b^{f})$$
(2)

3 Output Gate: controls the information of current cell state to make it visible and the equation is

$$o^{t} = \sigma(w^{o}(h^{t-1}, x^{t}) + b^{o})$$
(3)

All of the gates apply sigmoid activation function, because they constitute smooth curves in the range of 0 and 1 which makes model remains differentiable. The parameters w^i, w^f, w^o represent the weights at input, forget and output link respectively. h^{t-1} represents hidden state at time t - 1. Similarly b^i, b^f, b^o represents bias at input, forget and output link respectively, x^t represents input at time step t. Apart from these gates we have another vector C^t which modifies the cell state and is represented as:

$$C^{t} = \tanh(w^{c}(h^{t-1}, x^{t}) + b^{c})$$
(4)

The tanh activation helps cell state to avoid the vanishing gradient problem. In our work, the LSTM is trained to predict future Corona-Virus cases based on previously recorded real-world data. The model is implemented in Python (Pytorch) and trained on Google Colab (online GPU software). The data provided by John Hopkins University Center for Systems Science and Engineering (JHUC-SSE) *Github:https://github.com/CSSEGISandData/COVID-*19, which contains number of daily cases reported by each country. The sample of the data-set is given in figure 2. The data-set contains parameters like country, state, longitude, latitude which was not required for our model and were removed. Our model is trained on time-series data. The output at any time-step "t" depends on current input and previously generated output as shown in equation 1.

$$o_t = f(wo_{t-1} + wx_t) \tag{5}$$

here, $o_t \& o_{t-1}$ is output at time step t & t-1 respectively, x_t is the input at time step t. w is the weight associated to the connection link. f is the activation function. The dataset contains three sub-parts in which number of confirmed cases, recovered cases and death cases due to Covid-19 gets updated everyday (Globally). The architecture of our network is shown in figure 3. The LSTM is trained on



Touseef Iqbal, et al.: Deep Learning based Analysis of Covid-19 Mortality Risk

Province/State	Country/Region	Lat	Long	8/15/20	8/16/20	8/17/20	8/18/20	8/19/20	8/20/20	8/21/20	8/22/20	8/23/20
	Afghanistan	33.93911	67.709953	37551	37596	37599	37599	37599	37856	37894	37953	37999
	Albania	41.1533	20.1683	7260	7380	7499	7654	7812	7967	8119	8275	8427
	Algeria	28.0339	1.6596	38133	38583	39025	39444	39847	40258	40667	41068	41460
	Andorra	42.5063	1.5218	989	989	1005	1005	1024	1024	1045	1045	1045
	Angola	-11.2027	17.8739	1879	1906	1935	1966	2015	2044	2068	2134	2171
	Antigua and Barbuda	17.0608	-61.7964	93	93	93	93	94	94	94	94	94
	Argentina	-38.4161	-63.6167	289100	294569	299126	305966	312659	320884	329043	336802	342154
	Armenia	40.0691	45.0382	41495	41663	41701	41846	42056	42319	42477	42616	42792
Australian Capital Territory	Australia	-35.4735	149.0124	113	113	113	113	113	113	113	113	113
New South Wales	Australia	-33.8688	151.2093	3950	3957	3959	3966	3971	3972	3981	3985	3988

Figure 2. Data-set Sample



Figure 3. Architecture of LSTM network

all the three sub-parts of data-set. Unfortunately, we had data of only 212 days which is too little to train an LSTM model[17]. The data was split into two sections (test and train) 30 days of data for test-set and 182 days of data for training-set. Figure 4 & 5 shows training and test loss when LSTM is used for prediction of number of confirmed, recovered and death cases respectively. The graph was plotted between number of epochs and training/test loss for all the three cases using a python library Matplotlib. The graph initially shows higher training/test loss but slowly decreases as the model starts learning.

In the next attempt the whole data was used for training the network. Each training example contains sequence of data-points about history and a label of real value which our model predicts. To generate a chart for both historical (True cases) and predicted cases, the data-index was extended in the data frame for all the three sections of data-set. As shown in figure 6 & 7 the graph shows the comparison of predicted cases and the historical cases (True cases). The up-surge in predicted graph for number of patients, deaths and recovery shown in figures 6 & 7 may not seem convincing in comparison to the graph on true (historical) data. This is due to availability of less data, because deep learning models need lot of data to get trained. However, this is expected that the performance will increase once we have a large data-set to train the model.

The model was then trained to predict the cases on daily basis for next 12 days. Figure 8 & 9 shows the prediction for confirmed, recovered and death cases respectively. The more we train the network for future prediction, more we



5

Int. J. Com. Dig. Sys. 12, No.1, 1-8 (Jul-2022)



Figure 4. Loss function on death and recovered cases respectively



Figure 5. Comparison between predicted and true data on confirmed and recovered cases respectively.



Figure 6. Graph between predicted and true data on death cases

may not trust our model, that is due to less data. The use of generative modeling technique to enhance the data-set can be very helpful to make the predictions better. There have not been an excessive number of models and techniques that have firmly determined Corona-Virus spread. Deep learning models with attention mechanism, unsupervised learning and transfer learning can be very helpful to achieve better results in this area.

B. Generating the protein structure of Virus

Deep Learning can be helpful in number of ways. The diversity of models can be utilized at multistage of a problem to obtain extreme results. Another and very strong way to mitigate the issue of Corona-Virus disease 2019 is to generate the protein structure of the Virus. The protein structure explains the functioning of virus, which can be helpful for developing vaccines and drugs for it. Google Deep Mind has started this initiative towards this outbreak, trying to release protein structure of Covid-19 [18]. The diversity of models in deep learning like feature extraction (CNNs), Sequence modeling (RNNs) Generative models (GANs, VAEs) can be useful at any stage of problem. Based on the protein structure of previously recorded outbreaks like SARS-Cov-2 the protein structure of Covid-19 can be generated using any generative models of deep learning (GANs, VAEs). The generative adversarial networks can



Figure 7. Future prediction of confirmed and recovered cases respectively



Figure 8. Future prediction of Death cases



Figure 9. Loss function on confirmed cases

be trained to generate the protein structure of virus, to understand its functioning and to provide Vaccines and Drugs to stop this outbreak.

Deep learning can help us to slow down this outbreak and minimize the fatalities by slowing down the burden on the healthcare professionals and reminding patients about proper care procedures. The sharing of social data regarding movements of people and population density can be useful to predict the spread of the Virus.

4. DEEP LEARNING FOR COVID-19 DIAGNOSIS

Deep-Learning is also called as deep structured learning inspired with the structure of human brain, where layered architecture of model is used to analyze the data. Deep learning models uses the concept of representation learning "extraction of features from input data to define the problem" with multiple levels of representation. These representations consist of non-linear models which transform data from one level to another. The multiple level of representation helps to model the problems with complex nature. The diversity of applications areas demonstrates that deep learning can be applied in every domain. There are number of ways deep learning models can be used to deal with the problem of Covid-19 and similar outbreaks. In this section we will highlight how deep learning models can be used to tackle such problems.

A. Screening the Disease

Another methodology where deep learning can be exceptionally useful is at screening the disease. Diagnosing an instance of a Corona-Virus from CT (Computerized Tomography) output can conceivably reduce the time of diagnosis and facilitate better treatment. This approach can help us to calculate the probability of population positive with infection in any area. Also, figuring out the condition with the help of CT-outputs will improve the knowledge of sickness. The most common approach used here would be feature extraction method using Convolutional Neural Network (CNN). CNNs are class of neural networks inspired with the organization of visual cortex, where multiple



Int. J. Com. Dig. Sys. 12, No.1, 1-8 (Jul-2022)



Figure 10. Flowchart of Screening the disease

filters are used to extract the features critical to get good prediction. The feature extraction algorithm extracts the candidate regions of input image and these regions gets classified using any classification model and later infection probability is predicted. Similar work was proposed by [19] were deep neural networks with transfer learning based inception network was used. The pre-trained inception network converts the image data into one dimensional feature vectors and Convolutional Neural Network (CNN) was used for feature extraction. Bayesian function was applied to add uncertainty handling in the model. The image classification model was used to classify these features as relevant to infection or not. Lastly the probability of infection was determined using Noisy or Bayesian function. However the accuracy of the model can be improved by using algorithms with attention mechanism. The flowchart of this model is shown in figure 10.

5. CONCLUSION

In this paper we discussed some key concepts of Corona Virus disease 2019 (Covid-19). The approaches of deep learning which can be used for early prediction, screening and generating protein structure of the disease were discussed. In our work LSTM was trained on time series data to make prediction for Confirmed, Recovered and Death cases due to Covid-19. However, use of transfer learning, reinforcement learning and generative modeling would play an important role to counter Covid-19 and future pandemics. Since the data available is limited in terms of detailed patient history & symptoms etc. This research is in initial stage, there needs to be more work done on employing other techniques.

Deep Learning is still in its elementary stage. There is multiple range of opportunities for building, exploring current and new architectures to solve more complex problems. Deep learning models with local attention mechanism used for chest radiography can turn out to be more accurate for diagnosis of the disease.

References

- A. R. Fehr and S. Perlman, "Coronaviruses: an overview of their replication and pathogenesis," in *Coronaviruses*. Springer, 2015, pp. 1–23.
- [2] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in wuhan, china," *The Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.
- [3] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017.
- [4] F. He, Y. Deng, and W. Li, "Coronavirus disease 2019 (covid-19): What we know?" *Journal of Medical Virology*, 2020.
- [5] N. K. Karki, N. Verma, F. Trozzi, P. Tao, E. Kraka, and B. Zoltowski, "Predicting potential sars-cov-2 drugs-in depth drug database screening using deep neural network framework ssnet, classical virtual screening and docking," 2020.
- [6] A. Tampuu, Z. Bzhalava, J. Dillner, and R. Vicente, "Viraminer: Deep learning on raw dna sequences for identifying viral genomes in human samples," *PloS one*, vol. 14, no. 9, 2019.
- [7] S. Agrebi and A. Larbi, "Use of artificial intelligence in infectious diseases," in *Artificial Intelligence in Precision Health*. Elsevier, 2020, pp. 415–438.
- [8] Z. Bzhalava, A. Tampuu, P. Bała, R. Vicente, and J. Dillner, "Machine learning for detection of viral sequences in human metagenomic datasets," *BMC bioinformatics*, vol. 19, no. 1, p. 336, 2018.
- [9] V. Andriasyan, A. Yakimovich, F. Georgi, A. Petkidis, R. Witte, D. Puntener, and U. F. Greber, "Deep learning of virus infections reveals mechanics of lytic cells," *bioRxiv*, p. 798074, 2019.
- [10] F. Mock, A. Viehweger, E. Barth, and M. Marz, "Viral host prediction with deep learning," *bioRxiv*, p. 575571, 2019.
- [11] C. Shorten, T. M. Khoshgoftaar, and B. Furht, "Deep learning

applications for covid-19," Journal of Big Data, vol. 8, no. 1, pp. 1–54, 2021.

- [12] H. Li and F. Sun, "Comparative studies of alignment, alignment-free and svm based approaches for predicting the hosts of viruses based on viral sequences," *Scientific reports*, vol. 8, no. 1, pp. 1–9, 2018.
- [13] M. John and H. Shaiba, "Main factors influencing recovery in mers co-v patients using machine learning," *Journal of infection and public health*, vol. 12, no. 5, pp. 700–704, 2019.
- [14] L. Xu, R. Magar, and A. B. Farimani, "Forecasting covid-19 new cases using deep learning methods," *Computers in biology and medicine*, vol. 144, p. 105342, 2022.
- [15] I. Godfried, Machine Learning methods to aid in Coronavirus Response. Towards Data Science, 2020, no. 768.
- [16] R. Kremer, Using Kalman Filter to Predict Coronavirus Spread. Towards Data Science, 2020, no. 124.
- [17] V. Valkov, "Time series forecasting," no. 35, pp. 324-328, 2020.
- [18] M. AlQuraishi, "Alphafold at casp13," *Bioinformatics*, vol. 35, no. 22, pp. 4862–4865, 2019.
- [19] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Y. Chen, J. Su, G. Lang *et al.*, "Deep learning system to screen coronavirus disease 2019 pneumonia," *arXiv preprint arXiv:2002.09334*, 2020.

Touseef Iqbal Touseef Iqbal received B. Tech Degree in CSE from Islamic University of Science and Technology, India in year, 2015 and M .Tech Degree in CS from Central University of Jammu, India in year, 2018. He is currently a Research Scholar in the Department of CSE at National Institute of Technology, Srinagar, India. His research interests are in the field of Deep Learning, Unsupervised Learning and Deep Generative



Modeling



Dr. Shaima Qureshi Dr. Shaima Qureshi received her Ph.D. from National Institute of Technology Srinagar (NIT Srinagar). She has completed her B.E (Hons.) Computer Science degree from BITS Pilani, India in 2004. She completed her M.S in Computer Science from Syracuse University, NY, USA in 2006. She is working as an Assistant Professor in NIT Srinagar since 2008. She is currently guiding PhD, Masters and Bach-

elor of Engineering (B.E) students. She has 6 patents and 30 publications to her credit. Prior to joining the academic field, she worked as a Senior QA Engineer for two years in the software industry in USA. Her areas of research include Mobile Networks, Algorithms, Operating Systems and Computer Networks.