

An Analysis on Applications of Machine Learning Algorithms to predict COVID-19

Dr. A. P. Nirmala¹, Paramita Chandra²

¹Associate Professor

^{1,2}Department of MCA

New Horizon College of Engineering,

Bangalore, India

¹nirmalasuresh.ap@gmail.com, ²paramita.chandra98@gmail.com

Abstract

Machine Learning (ML) has proved itself as an important and distinguished field of study and research over the last decade by explaining and solving many very complex and sophisticated real-world problems. The application regions of ML included practically all this present reality spaces, for example, natural language processing, health care, business applications, autonomous vehicles, climate modelling, gaming, intelligent robots, image and voice processing. One of the most remarkable areas of ML is forecasting. There are loads of studies performed for the forecast of various diseases utilizing ML procedures such as cancer prediction, chronic disease prediction etc. Various standard ML calculations and algorithms have been utilized around there to direct the future course of activities and actions required in numerous applications including disease prediction, stock market prediction and weather forecasting etc. Different regression and neural system models have wide relevance in anticipating the states of patients later on with a particular illness. COVID-19 is presently considered as a potential threat to mankind. A few standard models for COVID-19 are being utilized by authorities around the world for enforcement of relevant control measures. To be explicit the examination will be focussed on the employments of ML on the COVID-19 episode. This study will focus on applications of ML on the healthcare industry pre-pandemic and how the existing methods can be enhanced and used for prediction and detection of COVID-19.

Keywords: Machine Learning; COVID-19; Prediction; Linear Regression (LR) models; Support Vector Machines (SVM); Multilayer Perceptron (MLP); Deep Learning.

1. Introduction

The COVID-19, an acronym for “Coronavirus Disease 2019”, is a severe acute respiratory syndrome coronavirus 2, also known as SARS-CoV-2 are a large family of viruses, including the Middle East Respiratory Syndrome (MERS)-CoV and SARS-CoV [1]. The International Health Regulations Emergency Committee of the World Health Organization (WHO) declared this new CoV called 'COVID-19', as pandemic and a 'general wellbeing crisis of global concern' on January 30, 2020. An accurate outbreak prediction model is required to obtain insights into the likely spread and consequences of these infectious diseases. Governments and other administrative and legislative bodies depend on bits of knowledge from prediction models to recommend new arrangements and to survey the viability of the implemented approaches and policies. The novel Coronavirus illness (COVID-19) has been accounted for to contaminate in excess of 2 million individuals, with in excess of 132,000 affirmed deaths around the world. The on-going worldwide COVID-19 pandemic has shown a nonlinear and complex nature. Furthermore, the episode has contrasts with other late flare-ups, which brings into question the capacity of standard models to convey precise outcomes.

This study plans to give an early estimate model to the spread of novel coronavirus. COVID-19 is presently an intense danger to human life everywhere throughout the world. To add to the current human emergency our endeavour in this investigation is to build up an estimating framework for COVID-19. ML works by perceiving plans in chronicled preparing data. The models dependent on Machine Learning have gotten so accommodative in their methodology that their presentation is getting more solid and sensibly precise with the range of time. ML calculations associated with cutting edge registering advancements support estimating limit and create versatility. ML has actually a wide extension when coordinated with business strategies. It very well may be of incredible use to stop various issues related with hierarchical usefulness. At present, most recent Machine Learning models can be useful to make forecasts extending from outbreaks of disease to the rise and fall of stocks. Machine Learning (ML) is two powerful tools or weapons that can be used to deal with this dynamic infectious disease.

The principle point of this review was to manage the association among viromics and Deep Learning and ML, which may amazingly help to discover an answer for this pandemic.

In this paper Section 2 represents the Literature Survey of the various ML techniques which were used for prediction and detection of diseases. Section 3 shows the application of machine learning algorithms to predict COVID-19, Section 4 presents the conclusion.

2. LITERATURE SURVEY

ML techniques were utilized in demonstrating previous pandemics (e.g., Ebola, Cholera, pig fever, H1N1 flu, and dengue fever, Zika, etc [6]. The ML strategies are restricted to the fundamental techniques for arbitrary woodland, neural systems, Bayesian systems, Naïve Bayes, genetic programming and Classification and regression tree (CART). In spite of the fact that ML has for quite some time been set up as a standard instrument for displaying catastrophic events and weather forecasting, its application in demonstrating episode is still in the beginning phases. More refined ML strategies (e.g., half and halves, troupes) are yet to be investigated. Consequently, the commitment of this paper is to investigate the utilization of ML for demonstrating the COVID-19 pandemic.

This paper plans to explore the speculation capacity of the proposed ML models and the precision of the proposed models for various lead-times. We give an extensive study of ML applications that would assist people to lessen and stifle the significant effects of the recent outbreak. There is a solid conviction that appropriate ML technique and research plans will fully exploit the intensity of ML in helping people to overcome the difficult fight against this deadly virus.

2.1 Effective Heart Disease Prediction using Hybrid Machine Learning Techniques

Heart disease is one of the most noteworthy reasons for mortality on the planet today. ML has been demonstrated to be successful in helping with settling on choices and expectations from the huge amount of information delivered by the health care industry [22]. Machine learning was also found to be effective in assisting in making useful decisions and predictions which were beneficial for the healthcare improvements. The proposed ML models and methods highlighted the features of machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. The prediction model based on ML introduced different combinations of features, and several known classification techniques. An enhanced ML model with high performance level and accuracy level of 88.7% through the prediction model for heart disease with Hybrid Random Forest with Linear Model (HRFLM) were introduced. There were various methods like K-Nearest Neighbor Algorithm (KNN), Decision Trees (DT), Genetic algorithm (GA), and Naive Bayes (NB) were also used [24] [27]. The model mainly focussed on UC Irvine (UCI) dataset classification and ML processes were used from a pre-processing data phase followed by feature selection based on DT entropy. The model proposed hybrid HRFLM approach which is used for combining the characteristics of Random Forest (RF) and Linear Method (LM) [26]. HRFLM model was proved to be quite accurate in the prediction of heart disease and maybe used for other diseases in near future with certain modifications.

2.2 Prediction of Cardiovascular Disease Risk for Newly Arrived Adult Canadian Immigrants Using Large Data and Machine Learning

This study aimed to evaluate the psychosocial hazard variables and resettlement stress connections to cardiovascular wellbeing among grown-up settler who arrived in Canada after 1985. Moreover, to create Machine Learning (ML) forecast models dependent on pre and present movement information on anticipate the danger of Cardiovascular Disease (CVD) for new appearances of grown-up workers as a component of the Watch Health Immigrant (WHI) Program [22][23]. It is mainly focused on the supervised machine learning, using an algorithm that iteratively learns from data of past events, and will be used to apply risk algorithm prediction models. Three algorithm prediction models were used here. The models were named as Population model, Individual model and Intervention models. The review companions were utilized to prepare and validate the prediction models. These models were helpful in preparing the prediction models, and were used for the estimation of individual risk and early intervention for the high-risk group of CVD for new arrivals on their first day in Canada based on pre-immigration and post-immigration data. It showed how ML supervised model could be more beneficial in prediction of CVD.

2.3 Role of biological Data Mining and Machine Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19)

The systematic review gave comprehensive overview of ML algorithms and data mining techniques with CoV family. CoV predictions algorithms were introduced. The evaluation conducted for each study included features like application nature; ML used and separated features and classes with accuracy percentages for the used ML algorithms were highlighted and demonstrated [28]. A lot of suggestions for the risk recovery of the Covid virus were established to fill in as a guide for future references in the field of data mining algorithms. In spite of the expanding paces of death and the quantity of individuals influenced with CoV datasets based on ML algorithms stays at a redefinition stage specifically for COVID-19. The analysis demonstrated numerous significant focuses. The analysis indicated many important points that are significant to identify the research gaps. In three ML techniques were applied to the MERS-CoV dataset to recognize the best arrangement model for binary class and multiclass labels. The outcomes indicated that the k-NN classifier is the best model for the two-class problems, and the decision tree and Naïve Bayes are the best models for multiclass problems [28]. In two experiments are applied to the MERS-CoV disease dataset, and the decision tree classifier shows higher prediction capability than different models [30]. The experimental outcomes demonstrated that age and symptoms are the two prevailing aspects for the forecast model and that medicinal services staff are probably going to endure. In an emotional recognition system based on ML technique was proposed to understand human responses to a widespread epidemic of transferrable diseases like MERS [28][30]. The study was conducted utilizing a dataset gathered in Korea in 2015 and the outcomes demonstrated the effect of helping unreasonable frenzy in lessening disease. In the creator researched individuals more than 50 years of age utilizing three ML techniques to anticipate MERS CoV [30][31]. The outcomes indicated that old individuals are more likely to be infected than others. In a Support Vector Machine (SVM) classifier dependent on sigmoid, ordinary and polynomial iterations was used to examine the MERS and SARS proteins [1]. The outcomes indicated their conduct closeness and inexact dissimilarities. In data mining based on statistical techniques was used to build up a cloud-based clinical framework or medical system with a high prediction accuracy to prevent MERS-CoV spread inside various districts [29]. The dataset includes the following characteristics: drug, patient and cloud-based user medical record.

This study intended to alleviate some of the challenges that have been addressed in the academic literature with their recommended solution for future studies. Studies, such as, suggested a pre-processing method to solve the missing-esteem issue that directly impacts the classification accuracy for the MERS infection. They likewise recommended the utilization of an ensemble technique by combining the cosine method with k-NN ML algorithm to improve the classification accuracy to >50%. Another study recommended increasing the number of samples by expanding the quantity of tests for the CoV dataset and gathering

data from patients within the same geographical area by straightforwardly speaking with devoted medical clinics and health organizations. The use of SVM classifier for the binary class problem and conducted an empirical study for multiclass problems in MERS-CoV. The utilization of the R language as a result of its effectiveness and high usefulness in supporting ML algorithms that can empower the improvement of an effective intelligent medical diagnosis system framework for CoV. Another study indicated the uncommon clinical issues identified with the female status, such as whether she is pregnant, which should be considered in the treatment of CoV-contaminated patients.

The Internet of Things (IoT) technology is found to be very useful for building up a profoundly reliable and highly dependable medical diagnosis system framework for COVID-19 [25].

2.4 A machine learning forecasting model for COVID-19 pandemic in India

The proposed model could be valuable by utilizing linear regression, Multilayer perceptron and Vector auto regression model on the COVID-19 kaggle data to anticipate the spread of COVID-19 cases in India. In 2020 a dynamic SEIR model for foreseeing the COVID-19 cases was introduced [7]. An ML model was used with respect to past SARS dataset moreover shows guarantee for future desire for the scourges. Another model was introduced at the beginning phase area of COVID-19, which is named by World Health Organization (WHO), by ML techniques actualized on stomach Computed Tomography pictures [8]. Another model presented presents a correlation of day level guaging models on COVID-19 influenced cases using time series models and numerical detailing.

In measurements, Linear Regression (LR) is a direct method to manage exhibiting the association between a dependent variable and at least one independent variable [10]. LR was the principle sort of regression analysis to be focused completely and to be used generally in valuable applications. LR shows the association between two factors by fitting a straight condition to based data [9][10]. A multilayer perceptron (MLP) is a type of feed forward artificial neural network (FANN). The term MLP is used every so often openly to show any FANN, from time to time cautiously to suggest frameworks made out of different layers of the perceptron [21][22]. An MLP is a perceptron that is generally used for complex issues. The structure of data based on date, confirmed, recovered and deaths were shown with the help of graphs and boxplots.

2.5 ML to accelerate the screening of COVID-19

ML techniques can be used to broaden screening and selection process by simultaneously considering several potential antiviral agents, relying on DNA sequences or protein structure including potential drug binding sites of the virus (SARS-CoV-2) to predict and analyse the interactions between drugs and the virus and can be efficiently used for shortlisting the promising individual treatments. ML has been previously used for other infectious diseases in a similar manner. For example: - Deep neural network was successfully trained to screen the activity of more than 100 million molecules on Escherichia Coli. Similarly, an enormous range of immunization competitors could be screened dependent on their capability to inspire a viable insusceptible reaction, for instance, by introducing the spike protein S that follows a SARS-CoV-2 infection [19]. Regardless, these possibly productive roads ought not to conceal the difficulties of helpful exploration dependent on ML. To begin with, ML can't quicken essential science, and even the expectation of protein collapsing stays an astoundingly troublesome problem. On account of immunizations, there is accordingly an essential holding up period [20]. Second, a significant moral concern is the impulse to sidestep appropriate clinical preliminaries working with very small cohorts, not using adequate design, or omitting inclusion and exclusion criteria have already be reported in the recent hydroxychloroquine-based treatment research. This risk could dramatically increase with ML algorithms. Indeed, algorithms such as deep neural networks are 'general approximators' [21]. Without a doubt they can be trained to fit any objective on a dataset such as remembering the determination of each patient. ML calculations must be assessed indisputably by surveying their capacity to precisely foresee an autonomous test set — a methodology that require enormous datasets and from the

earlier consideration and prohibition standards. The next section presents the different ML algorithms which can be used for accurate and successful detection of COVID-19.

3.APPLICATION OF MACHINE LEARNING ALGORITHMS TO PREDICT COVID-19

An ML based prediction system has been introduced for understanding and predicting the rise and risk of COVID-19 outbreak globally. The existing systems analyses datasets containing the day wise real past information and make prediction for upcoming days utilizing ML based systems. The most challenging tasks for these ML researchers is the lack of large datasets. The size of the data presently available is huge and so gathering information pattern out of it is really difficult and challenging as it has chances to hinder the understanding of the patterns and features of the virus. Therefore, the demand to construct a data set that can be understood and analysed the existing ML algorithms and techniques has increased since the current data sets includes info-graphic data and analysis of such kind of data is more difficult as compared to others. These models can be enhanced using several attributes like some deep learning methods for forecasting time series data for getting better results or adding factors like several hospitals, the immune system of the infected persons, or age and gender of the patients etc. Most of the existing models have shown to tend back to predictive applications of ML but it can be also used as descriptive and prescriptive applications with the help of Big data and different frameworks. This section presents the different algorithms of ML to predict and forecasting the Covid-19 outbreak.

3.1 Supervised Machine Learning Models

A supervised learning model is made to make a forecast when it is furnished with an unknown input. Accordingly, in this learning method, the learning algorithm takes a dataset with input occurrences alongside their corresponding regression to prepare and train the regression model. The prepared model at that point produces an expectation for the given unanticipated test data set or input data sets [2]. This learning strategy may utilize regression methods and order calculations for predictive model's advancement. Four regression models have been utilized in this investigation of COVID-19 future determining:

- Linear Regression (LR)
- Least Absolute Shrinkage and Selection Operator (LASSO) Regression
- Support Vector Machine (SVM)
- Exponential Smoothing (ES)

In Linear regression modelling, an objective class is predicated on the independent highlights and features [34].

This strategy can be in this manner used to discover the connection or relationship between independent and dependent variables and also for forecasting. To place the idea of linear regression in the Machine learning and setting, so as to prepare the model where x is represented as input training dataset, y represents the class labels present in the input dataset [2]. The objective of the ML learning algorithms and techniques at that point is to locate the best qualities for β_0 (capture) and β_1 (coefficient) to get the best-fit regression line. To get the best fit suggests the distinction between the actual values and predicted values should be reduced.

LASSO is a regression model belongs to the linear regression technique which uses shrinkage [5]. Here, Shrinkage refers to the shrinking of extreme values of a data set towards the central values. Thus, LASSO is better and more stable and also reduces the errors. For multi collinearity scenarios LASSO is considered as the best model.

A SVM is a kind of supervised ML algorithm utilized for both regression and classification [3][32]. SVM regression being a non-parametric strategy relies upon a lot of numerical capacities and set of mathematical functions. The set of functions known as kernel transforms the data inputs into the desired form. SVM solves the regression problems using a linear function, so while dealing with problems of

non-linear regression, it maps the input vector(x) to n -dimensional space called a feature space (z). This mapping is done by non-linear mapping techniques and after that linear regression is applied to space. Placing the idea of ML concepts with a multivariate training dataset (x_n) with N number of observations with y_n as a set of observed responses. The aim is to make it as level as conceivable subsequently to discover the estimation of $f(x)$ with $(\beta_0 \beta)$ as minimal norm values.

In exponential smoothing (ES) family methods, prediction is done based on previous period's data. The previous information perceptions impact is decaying exponentially as they become more seasoned. So, the weight assigned to different lag values is geometrically declined. ES is a very simple powerful time series forecasting method particularly for univariate data [4][33].

The study performs forecasts on death rate and as per results ES performs better among all the models, LR and LASSO perform equally well and accomplish nearly the equivalent to R^2 score but SVM performs worst in this circumstance. ES and LASSO lead in terms of performance for future forecast of new cases, LR also performed well, while SVM performs very poorly in terms of all the evaluation metrics. In recovery rate future forecasting the ES again performs better among all the other models. All other models perform poorly; the order of performance from best to worst is ES is best followed by LR, LASSO and SVM due to the nature of available time-series data. Thus, ES performs best in all three cases such as, death rate forecasting, the number of new confirmed cases forecasting, and recovery rate forecasting.

3.2 Multilayer Perceptron (MLP) and Adaptive neuro-fuzzy inference system or adaptive network-based fuzzy inference system (ANFIS) models

The worldwide pandemic of the serious intense respiratory disorder Coronavirus 2 (SARS-CoV-2) has become the essential national security issue of numerous countries. It is essential to advance the accurate prediction models for the outbreak to provide insights into the spread and consequences of this deadly disease. The standard epidemiological models for long-term prediction have shown low accuracy due to the lack of critical data and significant level of uncertainty. A comparative analysis of ML and soft computing models to predict the COVID-19 outbreak is shown. The outcomes of two ML models (MLP and ANFIS) reported a high speculation capacity for long-term predictions. The results reported in this study suggest ML as an effective tool to model the time series outbreak due to the highly complex nature of COVID-19 flare-up and contrasts from nation to nation.

The computer learns to use patterns or “training samples” in data (processed information) to predict or forecast and make intelligent decisions without over planning, using the ML methodologies [14][18]. In other words, ML is the scientific study of algorithms and statistical models used by computer systems that instead of using explicit instructions uses patterns and inference to perform tasks [15][16]. Time-series are data sequences collected over a period of time, which can be used as inputs to ML algorithms [17]. This type of data reflects the changes that a phenomenon has undergone over time. Let X_t be a time-series vector, in which x_t is the outbreak at time point t and T is the set of all equidistant time points. To train ML methods effectively, two different scenarios are used where scenario 1 employs data for three weeks to predict the outbreak on day t and scenario 2 employs outbreak data for five days to predict the outbreak for day t . Both of these scenarios were employed for fitting the ML methods. Two frequently used ML methods here are the multi-layered perceptron (MLP) and adaptive network-based fuzzy inference system (ANFIS) are employed for the prediction of the outbreak in the five countries.

3.3 Long Short-Term Model (LSTM) model

The Machine learning and Deep learning approaches output the possible no of cases for the next 10 days across the globe. The predicted trend of COVID -19 can be illustrated using Support Vector Regression (SVR), Deep neural networks (DNN), Convolutional Neural Network (CNN), LSTM and polynomial regression (PR) with worldwide data [13]. It is also observed that the training and preparation of LSTM model highly depends on the deviation in the values, and it is also concluded that larger the deviation the more time it would take to train the model. So, the no of cases should be scaled using minmax scaler so

that it would fit the LSTM model and further the predicted cases could also be rescaled to the original range using invert minmax transform from sklearn python library. PR approach is best for the growing trend. From the results it is found that the (PR) gave a minimum root mean square error (RMSE) score over other methods in prediction of COVID -19 transmissions [11][12]. There would be a huge loss of lives if it follows the predicted trend of the PR mode as it shows the exponential growth of transmission worldwide. This growth of COVID -19 can only be reduced by decreasing the no of suspect able individuals from the infected individuals. The study can further extend by utilizing other DL & ML methodologies and models.

3.4 Detection of X ray images with ML based tools.

ML based tools such as Tenserflow & Keras could possibly detect COVID-19 in X-ray images of patients. This method is not so reliable and not an accurate medical diagnostic method but it could still show the knowledge & power of machine learning and deep learning and how ML based systems could make a great impact on the world of healthcare. Applying ML into this would be very efficient as we just need to gather the X-ray dataset of patients who have been tested Covid positive and then it could be followed by sampling of normal patient. These data can be easily obtained from various open sources. At last a CNN should be trained to detect COVID -19 from the sampled X-ray images. Neural networks could also be used to train these models for automatic detection of the presents of COVID-19.

3.5 Exploring SARS-CoV-2 Spike Glycoprotein with python tools

Different NCBI viruses' database can be used to download beta corona virus sequence and then subset database sequence can be obtained. These subsets could be used to annotate the nucleotide sequences and it can be saved in a data frame. These sequences could be aligned and accessed from the data frames. By this the behaviour of the sequence could be noted and can be used further for ML tools and frameworks. Here different neural networks concepts of deep learning could also be included.

Table 1. COMPARISON OF DIFFERENT ML TECHNIQUES FOR FUTURE FORECASTING AND PREDICTION OF COVID

ML Techniques	Type	Accuracy level and chances of error in prediction	Performance with respect to small data sets	Performance with respect to large data sets	Efficiency for different conditions
LASSO	Linear Regression	Average	Very poor	Good	Most efficient for future forecasting of new infected confirmed cases
SVM	Regression and Classification	Low	Very poor	Average	Less efficient for all the cases
LR	Statistical technique for predictive analysis	Average	Very poor	Best	Most efficient for forecasting uncertainty

ES	Smoothing family methods	High	Best	Best	Most efficient for future forecasting for all the cases
MLP	Based on neural networks	High	Best	Good	Efficient for long term prediction
ANFIS	Fuzzy based structures	High	Best	Average	Efficient for long term prediction

The above table shows a comparison of different ML techniques and their accuracy, efficiency and performance level in prediction and future forecasting of COVID with respect to small and large data sets.

4. CONCLUSION

In this study, various existing Machine Learning models and tools which can be used for the prediction or can be helpful for the fight against the COVID-19 virus are discussed in depth. The methodologies which can be used for further enhancement and development of new ML based tools are also mentioned. It is shown that some of the Deep Learning, Neural Network and Big Data concept can also be included for increasing the efficiency of existing models or systems. The challenge for large scale implementation of the models is the huge data requirement. This should be overcome and various attributes which could increase the quality of data acquired should also be included and improved. The study has sufficient information regarding the gaps of the existing models and has suggestions for improvements and new models.

5. REFERENCES

5.1 Journal Article

- [1] Al-Turaiki, M. Alshahrani, and T. Almutairi, “Building predictive models for MERS-CoV infections using data mining techniques,” *J. Infect. Public Health*, vol. 9, no. 6, pp. 744–748, 2016.
- [2] Grushka-Cockayne and V. R. R. Jose, “Combining prediction intervals in the m4 competition,” *Int. J. Forecasting*, vol. 36, no. 1, pp. 178–185, Jan. 2020.
- [3] X. F. Du, S. C. H. Leung, J. L. Zhang, and K. K. Lai, “Demand forecasting of perishable farm products using support vector machine,” *Int. J. Syst. Sci.*, vol. 44, no. 3, pp. 556–567, Mar. 2013.
- [4] E. Cadenas, O. A. Jaramillo, and W. Rivera, “Analysis and forecasting of wind velocity in chetumal, quintana roo, using the single exponential smoothing method,” *Renew. Energy*, vol. 35, no. 5, pp. 925–930, May 2010.
- [5] K. M. Anderson, P. M. Odell, P. W. Wilson, and W. B. Kannel, “Cardiovascular disease risk profiles,” *Amer. heart J.*, vol. 121, no. 1, pp. 293–298, 1991.
- [6] H. Asri, H. Mousannif, H. A. Moatassime, and T. Noel, “Using machine learning algorithms for breast cancer risk prediction and diagnosis,” *Procedia Comput. Sci.*, vol. 83, pp. 1064–1069, Jan. 2016.
- [7] Sujatha R, Chatterjee J (2020) A machine learning methodology for forecasting of the COVID-19 cases in India.
- [8] Yan X, Su X (2009) Linear regression analysis: theory and computing. World Scientific, New York.
- [9] Barstugan M, Ozkaya U, Ozturk S (2020) Coronavirus (covid-19) classification using ct images by machine learning methods. arXiv preprint arXiv:2003.09424.
- [10] Rizk-Allah RM, Hassanien AE (2020) COVID-19 forecasting based on an improved interior search algorithm and multi-layer feed forward neural network. arXiv preprint arXiv:2004.05960.
- [11] Elmousalami HH, Hassanien AE (2020) Day level forecasting for Coronavirus Disease (COVID-19) spread: analysis, modeling and recommendations. arXiv preprint arXiv:2003.07778.

- [12] Remuzzi, A.; Remuzzi, G. COVID-19 and Italy: what next? *Lancet* 2020.
- [13] Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey wolf optimizer. *Advances in engineering software* 2014, 69, 46– 61.
- [14] Ardabili, S.; Mosavi, A.; Varkonyi-Koczy, A. Building Energy information: demand and consumption prediction with Machine Learning models for sustainable and smart cities. *Engineering for Sustainable Future, Lecture Notes in Networks and Systems*, Springer Nature Switzerland.: 2019.
- [15] Gundoshmian, T.M.; Ardabili, S.; Mosavi, A.; Varkonyi-Koczy, A.R. Prediction of combine harvester performance using hybrid machine learning modeling and response surface methodology. 2019.
- [16] Nosratabadi, S., Mosavi, A., Duan, P. and Ghamisi, P., 2020. Data Science in Economics. *arXiv preprint arXiv:2003.13422*.
- [17] Hamilton, J.D. Time series analysis; Princeton university press Princeton, NJ: 1994; Vol. 2.implementations
- [18] Rodriguez-Morales AJ, Bonilla-Aldana DK, Balbin-Ramon GJ, Rabaan AA, Sah R, Paniz-Mondolfi A, et al. History is repeating itself: Probable zoonotic spillover as the cause of the 2019 novel Coronavirus Epidemic. *Infez Med.* 2020;28(1):3–5. [PubMed: 32009128].
- [19] Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmun.* 2020;109:102433. doi: 10.1016/j.jaut.2020.102433. [PubMed: 32113704]. [PubMed Central: PMC7127067].
- [20] Sundaram L, Gao H, Padigepati SR, McRae JF, Li Y, Kosmicki JA, et al. Predicting the clinical impact of human mutation with deep neural networks. *Nat Genet.* 2018;50(8):1161–70. doi: 10.1038/s41588-018-0167-z. [PubMed: 30038395]. [PubMed Central: PMC6237276]. 15. Werbos PJ. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE.* 1990;78(10):1550–60. doi: 10.1109/5.58337.
- [21] Basheer IA, Hajmeer M. Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods.* 2000;43(1):3–31. doi: 10.1016/s0167-7012(00)00201-3.
- [22] Durairaj, M. Revathi, V., 2015. Prediction Of Heart Disease Using Back Propagation MLP Algorithm. , 4(08), pp.235–239.
- [23] Gavhane, A., 2018. Prediction of Heart Disease Using Machine Learning. 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), (Iceca), pp.1275–1278.
- [24] Gushulak BD, Williams LS. National Immigration Health Policy Existing Policy, Changing Needs, and Future Directions. 2004.
- [25] Dr. A P Nirmala, Suchandra Paul, Subham Mahapatra Internet of Things Based on Smart Agriculture System Using Predictive Analysis, *International Journal of Recent Trends in Engineering & Research (IJRTER)* Volume 04, Issue 03; March- 2018 [ISSN: 2455-1457].

5.2 Book

- [26] Dunn JR, Dyck I. Social determinants of health in Canada’s immigrant population: results from the National Population Health Survey, 2000.
- [27] Canadian Mental Health Association. Immigrant and Refugee Mental Health: Backgrounder – Canadian Mental Health Association Canadian Mental Health Association. WDgTJvkrLIV. Published 2004.

5.3 Chapter in a Book

- [28] N. Almansour, H. Kurdia, N. Almansour, and H. Kurdia, “Identifying accurate classifier models for a text-based MERSCoV dataset,” in 2017 Intelligent Systems Conference, IntelliSys 2017, 2018, vol. 2018-Janua, pp. 430–435.

- [29] S. Choi, J. Lee, M. G. Kang, H. Min, Y. S. Chang, and S. Yoon, “Large-scale machine learning of media outlets for understanding public reactions to nation-wide viral infection outbreaks,” *Methods*, vol. 129, pp. 50–59, 2017.
- [30] M. Abdullah, M. S. Altheyab, A. M. A. Lattas, and W. F. Algashmari, “MERS-CoV disease estimation (MDE) A study to estimate a MERS-CoV by classification algorithms,” in *Communication, Management and Information Technology - Proceedings of the International Conference on Communication, Management and Information Technology, ICCMIT 2016, 2017*, pp. 633–638.
- [31] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009.
- [32] F. Rustam, I. Ashraf, A. Mehmood, S. Ullah, and G. Choi, “Tweets classification on the base of sentiments for US airline companies,” *Entropy*, vol. 21, no. 11, p. 1078, Nov. 2019.
- [33] F. Petropoulos and S. Makridakis, “Forecasting the novel coronavirus COVID-19,” *PLoS ONE*, vol. 15, no. 3, Mar. 2020, Art. no. e0231236.
- [34] P. Lapuerta, S. P. Azen, and L. Labree, “Use of neural networks in predicting the risk of coronary artery disease,” *Comput. Biomed. Res.*, vol. 28, no. 1, pp. 38–52, Feb. 1995.