Forecasting SARS-CoV-2 Next Wave in Iran and the World Using TSK Fuzzy Neural Networks

Seyed Muhammad Hossein Mousavi [1,*] - S. Muhammad Hassan Mosavi [2]
[1] Dept of Computer Engineering, Bu Ali Sina University
[2] Department of Art, University of Applied Sciences

ABSTRACT

Based on World Health Organization (WHO) report, 464,596 confirmed, 26,567 death and 392,293 recovered cases of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) or COVID-19 are reported as of October 6, 2020 for Iran. This virus became pandemic in March 11, 2019 and spread worldwide. Due to the absence of specific vaccine, non-pharmacological interventions like social distancing, using disinfectants and wearable masks and gloves are essential till finding an absolute solution. But during this time, Artificial Intelligence methods could aid the available methods to reduce the number of confirmed cases by forecasting the future based on available data from people of a specific region. Fortunately, there are considerable number of up to data datasets of COVID-19 which contain time-series data. Three best of these datasets are John Hopkins University, WHO and European Centre for Disease Prevention and Control datasets which this paper is employed them for experiment. This paper uses Takagi Sugeno Kang Fuzzy Neural Networks forecasting system on mentioned datasets for duration between two waves which is around 3.5 months. Some of the validation metrices are Mean Absolute Percentage Error (MAPE), Explained Variance (EV), and Root Mean Squared Log Error (RMSLE). By forecasting the first wave data for the second wave, highly similar to second wave data are forecasted by the proposed system which, makes this system robust and applicable for future waves.

Keywords: *COVID-19, Forecasting, World Health Organization, Time-Series, Fuzzy Neural Networks*

1. INTRODUCTION

The starting point of this epidemic was in December 2019, at Hubei province and its capital, Wuhan. As of today (October 6, 2020), we are in the middle of third wave of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) or COVID-19 in Iran and most parts of the world. 34,396,222 people are infected by this virus worldwide [1] so far and increasing. However, Iran's share is 464,596 [1]. These claims are based on World Health Organization (WHO) report [1]. It is not guaranteed that the proper vaccine made before the fourth wave, so forecasting the future is a rational action. There is a note should be considered about forecasting and prediction. A forecast refers to a calculation or an estimation which uses data from previous events, combined with recent trends to come up a future event outcome. Forecast implies time series and future, while prediction does not. A prediction is a statement which tries to explain a possible outcome or future event. The paper is dealing with data and it is all about forecasting of number of confirmed, death and recovered cases. To know this virus better, Figure 1 could provide enough details of its structure. This virus is made of 6 main parts of Nucleocapsid protein and RNA, Spike glycoprotein, Lipid bilayer membrane, Hemagglutinin esterase, Membrane protein and Envelope protein which makes this virus so dangerous [2]. Also, Figure 2 represents microscopic image of SARS-CoV-2 cells in act.

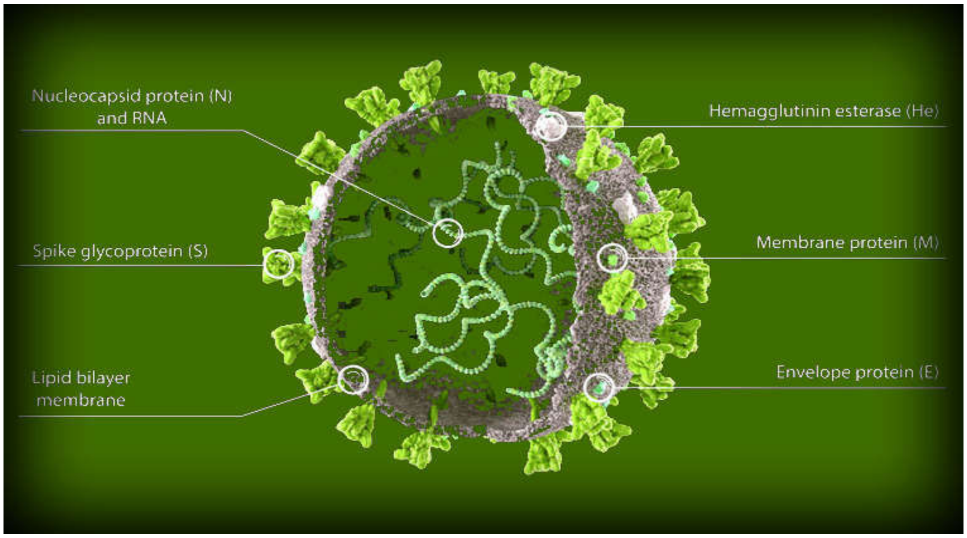

Fig. 1. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) or COVID-19 cell structure

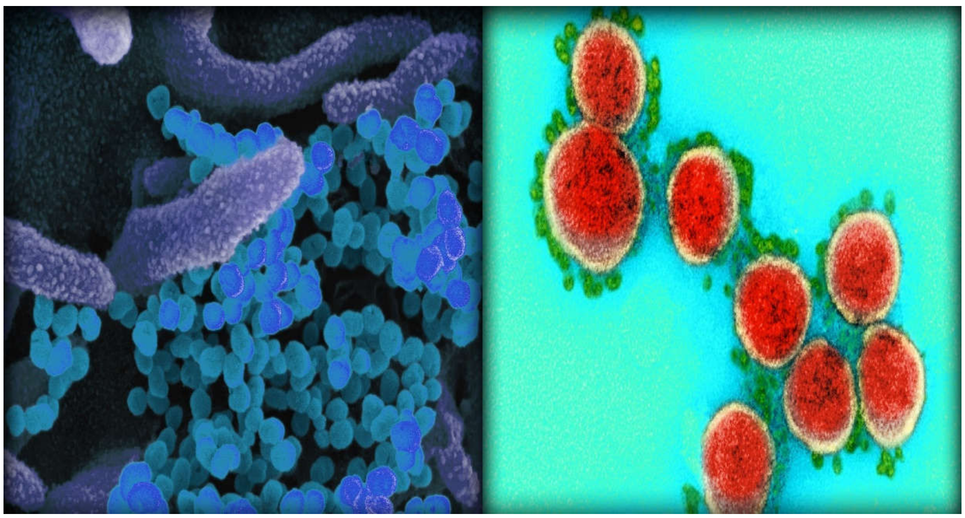

Fig. 2. Microscopic image of SARS-CoV-2 cells [3]

There is branch of Artifactual Intelligence called Datamining [4, 5]. Available datasets are in the form of numerical matrixes. Through machine learning [6] algorithms, it is possible to mine and learn these datasets and produce new meaningful information. This paper aims to use Neural Networks [7] alongside with Fuzzy logic [8] to forecast the future of this epidemy in Iran and the world. The experiment algorithm is called Takagi Sugeno Kang Fuzzy Neural Networks [9, 15] which is very robust and still is not used for forecasting COVID-19 yet. The performance of this algorithm during experiment and on this subject was significant, so it is decided to employ it.

1.1 Machine Learning and Datamining

Machine learning means that a computer program can learn and adapt to new position without human intervention. It is a field of artificial intelligence (AI) that keeps a computer's built-in algorithms current regardless of changes in the world. In the other hand, Data mining is the process of finding specific, patterns and correlations within large amount of data to predict results. Using a broad range of techniques, you can use this information to increase efficiency, decrease costs, improve customer relationships, reduce risks and more [10, 11].

1.2 Neural Networks and Fuzzy Logic

A neural network [12] is a series of algorithms that tries to recognize hidden relationships in a dataset through a process that simulate the way the human brain works. In order to do that, neural networks refer to systems of neurons, either organic or artificial in the nature.

Fuzzy logic means to computing based on "degrees of truth" rather than the usual "true or false" (1 or 0) or Boolean logic on which the modern computer is based. The idea of fuzzy logic was first made by Dr. Lotfi Zadeh of the University of California at Berkeley in the 1960s [12, 13]. A fuzzy system consists of four components: fuzzifier, rule base, inference engine, and defuzzifier. The fuzzifier maps each crisp input into a fuzzy set, the inference engine performs inferences on these fuzzy sets to obtain another fuzzy set, utilizing the rule base, and the defuzzifier converts the inferred fuzzy set into a crisp output. Figure 3 represents the fuzzy system structure.

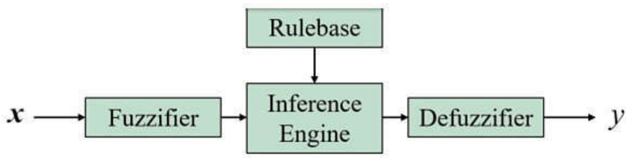

Fig. 3. Four main parts of a fuzzy system

The paper is consisting of 5 main sections. Section 1 described introduction and fundamentals of the research. Section 2 paid to Prior related works, done by other researchers on COVID-19 forecasting. Section 3 is all about proposed Adaptive-Network-based Fuzzy (TSK) Inference System (ANFIS) COVID-19 forecasting system, validations and final results and discussion about it. Conclusion, suggestion and future work concluded in section 4. Final section includes related citations for the research.

2. PRIOR RELATED RESEARCHES

In 2020, Roosa, Kimberlyn, et al, used daily reported cumulative case data up until 13 February 2020 from the National Health Commission of China. They report 5- and 10-day ahead forecasts of cumulative case reports. Specifically, they generate forecasts using a generalized logistic growth model [18]. They used National Health Commission of China dataset.

Next research belongs to Al-Qaness, Mohammed AA, et al in 2020. This research forecasts confirmed cases in China based on previously confirmed cases. They used evolutionary improved ANFIS to do this. They used WHO dataset for upcoming next 10 days. They achieved nice results but they used just one dataset and for the second wave. Also, they used traditional fuzzy system, not TSK. The other problem was that they forecast just next 10 days which is not practical. Also, they forecast just confirmed case, not death and recovered case, which proposed research fixed all these problems [19].

Another valuable research for China, is Roosa, K, et al forecast system for next 15 days, from February 5th to February 24th, 2020 [20]. They collected up to date data from National Health Commission of China website.

Also, a research belongs to Fanelli, Duccio, and Francesco Piazza in 2020 shows COVID-19 forecasting results for 22 / 01 −15 / 03 / 2020 time period for China, Italy and France. This research covered confirmed, recovered and death cases in these countries [21]. Data were gathered from the GitHub repository associated with the interactive dash- board hosted by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, Baltimore, USA.

One of the valuable researches for modeling and forecasting COVID-19 pandemic in India is belong to Sarkar, Kankan, Subhas Khajanchi, and Juan J. Nieto. They covered 17 provinces data from WHO report for India.

Ghanbari, Behzad made a forecasting system for the second wave of COVID-19 in Iran. Actually, his forecasting was on people with weak immune system [22]. People which had diseases such as heart disease, cancer or diabetes. Although his data was taken from Wikipedia, but the research is valuable. It has to mention that his data about people with weak immune system is taken from [23].

Kapoor, Amol, et al, had a valuable experiment on forecasting COVID-19 next wave using Spatio-Temporal Graph Neural Networks on US county level COVID-19 dataset, New York Times (NYT) COVID-19 dataset, the Google COVID-19 Aggregated Mobility Research Dataset and the Google Community Mobility Reports and achieved significant results [24].

Zeroual, Abdelhafid, et al, used five Deep learning algorithms to forecast new and recovered case of COVID-19. The data is based on daily confirmed and recovered cases collected from six countries namely Italy, Spain, France, China, USA, and Australia of WHO report [25].

3. PROPOSED FORCASTING SYSTEM AND VALIDATION

3.1 Takagi Sugeno Kang Fuzzy Neural Networks

There are two types of rules for a fuzzy system: Zadeh [14], where the rule consequents are fuzzy sets, and Takagi-Sugeno-Kang (TSK) [9, 15], where the rule consequents are functions of the inputs. SK rules are much more popular in practice due to their simplicity and flexibility. Researchers started to discover in the early 1990s that a TSK fuzzy system can be represented similarly to a neural network [16], so that a neural network learning algorithm, such as back-propagation [11], can be used to train it. These fuzzy systems are called neuro-fuzzy systems. The most popular one may be the Adaptive-Network-based Fuzzy Inference System (ANFIS) [17].

In Figure 4 (a), The first layer represents inputs, the middle (hidden) layer represents fuzzy rules, and the third layer represents outputs. Fuzzy sets are encoded as connection weights. Forecasting TSK fuzzy neural network ANFIS structure with five layers is presented in Figure 4 (b). First layer is for membership grades (confirm, death and recovered cases of COVID-19 datasets). After preparing data for time series, it will be split to train and test. Second layer belongs to firing level of each rule, the normalized firing level of the rules makes layer three, in layer four each normalized firing level is multiplied by its corresponding rule and finally there is output layer of forecast. All parameters of the ANFIS, the shapes of the Membership Functions (MFs (here, gaussian)) and the rule consequents, can be trained by a gradient descent algorithm [11].

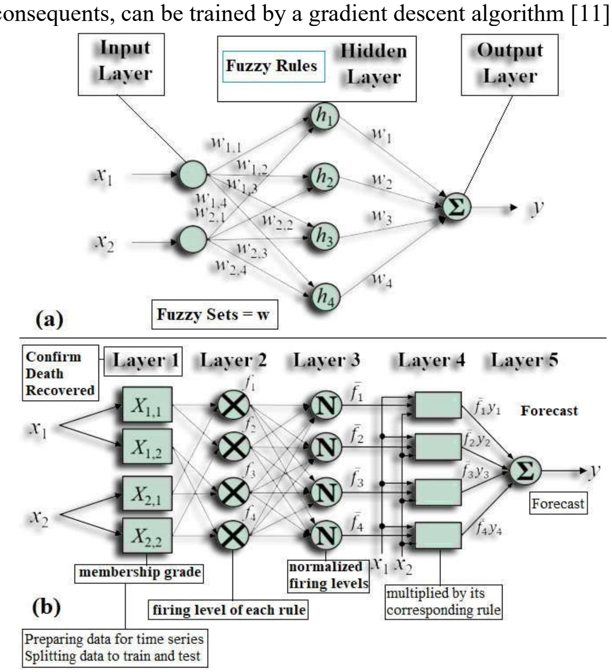

Fig. 4. (a). Inputs, the middle (hidden) layer or fuzzy rules and outputs, (b). The TSK fuzzy system, represented as a 5-layer ANFIS for forecasting

3.2 Datasets

Three well documented, most up to date, daily COVID-19 datasets which covers confirmed, death and recovered cases are Johns Hopkins University [27], World Health Organization (WHO) [1] and European Centre for Disease Prevention and Control [28] datasets which in this paper are employed. Except mentioned factors, datasets cover exact data for all waves, countries, continent, cumulative deaths and lots of other factors which are not related to the subject but useful for other researchers. For example, Johns Hopkins University dataset covers longitude and latitude of cases in cities, which helps for classifying the regions based on their level of infection.

3.3 Validation

Validation metrices are Root Mean Square Error (RMSE) (1) [25], Mean Absolute Percentage Error (MAPE) (3) [25], Mean Absolute Error (MAE) (2) [26], Explained Variance (EV) (4) [25], and Root Mean

Squared Log Error (RMSLE) (5) [25]. These five-performance metrics are used for three confirmed, death and recovered factors and for three mentioned datasets to forecast the next wave.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_l - \hat{y_t})^2} \quad (1)$$

$$MAE = \frac{\sum_{t=1}^{n}|y_t = \hat{y_t}|}{n} \quad (2)$$

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_l - \hat{y_t}}{y_l}\right|\% \quad (3)$$

$$EV = 1 - \frac{Var(\hat{y} - y)}{Var(y)} \quad (4)$$

$$RMSLE_{=}\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\log(y_t) - \log(\hat{y_t}))^2} \quad (5)$$

where y_l are the actual values, $\hat{y_t}$ are the corresponding estimated values, and n is the number of measurements. The benefit of using RMSLE as the statistical indicator is that its great robustness to outliers. Lower RMSE, MAE, or MAPE values and EV closer to 1 represent more accurate forecasting performances. Employed data for all datasets belong to 281 days or 9 months and 36 days till October 6, 2020. But John Hopkins University covered 258 days. These data are divided to 75 % for train and remaining 25 % for test purposes. Figure 5 shows confirm cases forecasting result by proposed method for Iran data. It shows increasing values.

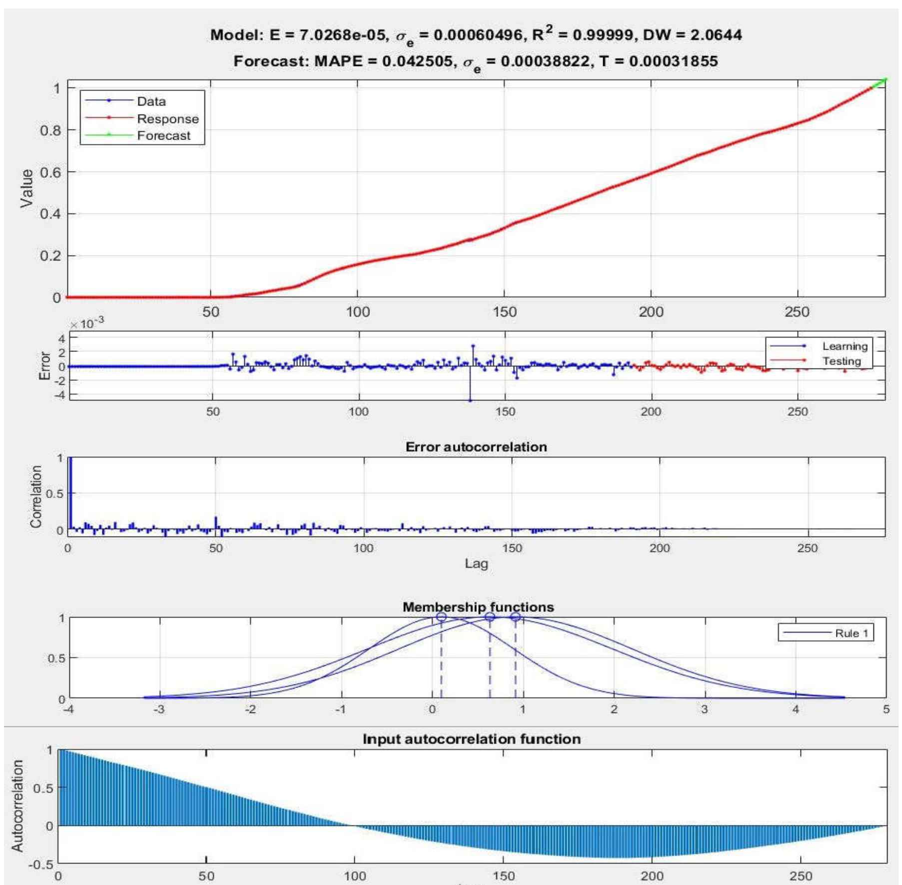

Fig. 5. Confirm cases forecasting for Iran data by proposed system up to two months.

Table 1 represents acquired results for confirmed cases in Iran and the world using proposed approach for five mentioned performance metrics and comparison with [25] (RNN) paper. Also, Tables 2 and 3 represent acquired results for death and recovered cases with same features as for Table 1.

Table 1. Validation Metrics for Confirmed Cases of COVID-19 Forecasting Using TSK ANFIS and Comparison Results with [25] (RNN) For All Datasets

Region	Model	RMSE	MAE	MAPE	EV	RMSLE
Iran	Proposed	1,112	1,136	3956	0709	00021
	[25] (RNN)	1,080	1,122	4012	0302	00044
World	Proposed	2,111	2,971	8363	0199	00782
	[25] (RNN)	2,980	3,871	8471	0097	00911

Table 2. Validation Metrics for Death Cases of COVID-19 Forecasting Using TSK ANFIS and Comparison Results with [25] (RNN) For All Datasets

Region	Model	RMSE	MAE	MAPE	EV	RMSLE
Iran	Proposed	1,342	1,655	4333	0607	00111
	[25] (RNN)	1,118	1,529	4187	0412	00188
World	Proposed	3,218	3,811	9363	0223	00606
	[25] (RNN)	3,360	4,011	9981	0088	00866

Table 3. Validation Metrics for Recovered Cases of COVID-19 Forecasting Using TSK ANFIS and Comparison Results with [25] (RNN) For All Datasets

Region	Model	RMSE	MAE	MAPE	EV	RMSLE
Iran	Proposed	1,022	1,001	2601	0912	00007
	[25] (RNN)	1,037	1,008	3013	0816	00019
World	Proposed	1,196	1,288	6500	0414	00099
	[25] (RNN)	2,000	2,117	7018	0400	00126

3.4 Discussion

One of the best researches on this subject was [25]. So, it was decided to use this paper's implement for comparison purpose. They used five different deep learning approached which RNN had the best result, so this approach is selected for comparison. As it is clear in Tables I, II and III, proposed method could achieve promising results for different performance metrics and in comparison, with [25] method. Figure 6 presents acquired results using proposed TSK ANFIS forecasting system for confirmed, death and recovered variables which belongs to Iran data achieved from all three datasets. The same details for the world are presented in Figure 7. It was decided to forecast next 90 days or three months from Oct 6 2020 till Jan 6 2021. Forecasting results shows stable values for all three variables in duration of first 28 days and increasing for next 32 days and decreasing values for last 30 days. System forecasts, fourth was arrives in Dec 6 2020 and it is more powerful than preview waves. If a proper vaccine does not distribute globally, based on our system, at the end of 90 days, 70000000 million people will be infected by this deadly virus. But if no vaccine distributes, there is enough time after Nov 8 2020 till Dec 6 2020 to increase non-pharmacological interventions policy worldwide.

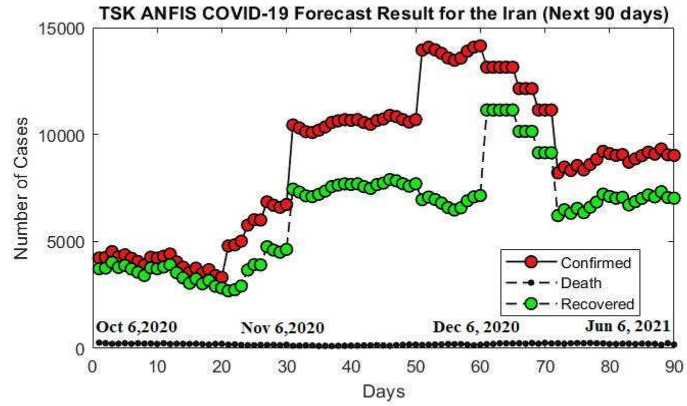

Fig. 6. TSK ANFIS Forcasting Results for Iran

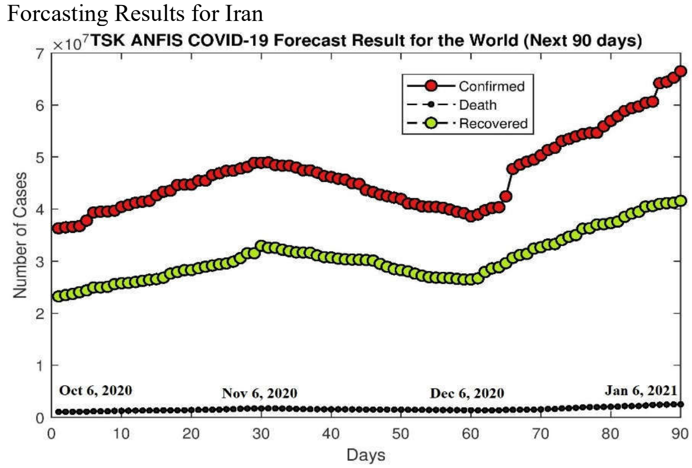

Fig. 7. TSK ANFIS Forcasting Results for the World

4. CONCLUSION, SUGGESTION AND FUTURE WORK

As COVID-19 has more destructive effects than other corona type viruses like MERS and SARS and no proper vaccine is made, forecasting the outcome is essential to use proper prevention methods instead. With combining artifactual intelligence techniques, such as fuzzy logic, machine learning and neural networks, it is possible to achieve the higher quality results and for longest period of time which has higher precision. The lack of long period predictions in available researchers for Iran, made this experiment possible to fill the gap, as this research covered up to three months of forecasting. Most of the researches drops recovered cases, which was one of the aims of the research. It is suggested to apply this method on other available datasets downloadable from [29]'s links. This research found gaussian membership function useful for used datasets. Other types of membership functions may work better for other data. Also, increasing performance metrics and forecasting reinfection of recovered cases is of future works.

REFERENCES

[1] https://covid19.who.int/
[2] Mousavizadeh, Leila, and Sorayya Ghasemi. "Genotype and phenotype of COVID-19: Their roles in

pathogenesis." Journal of Microbiology, Immunology and Infection (2020).
[3] https://www.niaid.nih.gov/news-events/novel-coronavirus-sarscov2-images
[4] Mousavi, Seyed Muhammad Hossein, Vincent Charles, and Tatiana Gherman. "An evolutionary Pentagon Support Vector finder method." Expert Systems with Applications 150 (2020): 113284.
[5] Mousavi, Seyed Muhammad Hossein, S. Younes MiriNezhad, and Atiye Mirmoini. "A new support vector finder method, based on triangular calculations and K-means clustering." 2017 9th International Conference on Information and Knowledge Technology (IKT). IEEE, 2017.
[6] Dezfoulian, Mir Hossein, et al. "Optimization of the Ho-Kashyap classification algorithm using appropriate learning samples." 2016 Eighth International Conference on Information and Knowledge Technology (IKT). IEEE, 2016.
[7] Mousavi, Seyed Muhammad Hossein, and VB Surya Prasath. "Persian Classical Music Instrument Recognition (PCMIR) Using a Novel Persian Music Database." 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE). IEEE, 2019.
[8] Mousavi, Seyed Muhammad Hossein, et al. "A PSO fuzzy-expert system: As an assistant for specifying the acceptance by NOET measures, at PH. D level." 2017 Artificial Intelligence and Signal Processing Conference (AISP). IEEE, 2017.
[9] Zhang, Yunlong, and Hancheng Ge. "Freeway travel time prediction using Takagi–Sugeno–Kang fuzzy neural network." Computer-Aided Civil and Infrastructure Engineering 28.8 (2013): 594-603.
[10] Russell, Stuart, and Peter Norvig. "Artificial intelligence: a modern approach." (2002).
[11] Duda, Richard O., Peter E. Hart, and David G. Stork. Pattern classification. John Wiley & Sons, 2012.
[12] Nielsen, Michael A. Neural networks and deep learning. Vol. 2018. San Francisco, CA: Determination press, 2015.
[13] Zadeh, Lotfi Asker, George J. Klir, and Bo Yuan. Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers. Vol. 6. World Scientific, 1996.
[14] S. S. L. Chang and L. A. Zadeh, “On fuzzy mapping and control,” IEEE Trans. on Systems, Man, and Cybernetics, vol. SMC-2, no. 1, pp. 30–34, 1972.
[15] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its application to modeling and control,” IEEE Trans. on Systems, Man, and Cybernetics, vol. 15, pp. 116–132, 1985.
[16] S. K. Halgamuge and M. Glesner, “Neural networks in designing fuzzy systems for real world applications,” Fuzzy Sets and Systems, vol. 65, no. 1, pp. 1–12, 1994.
[17] J. R. Jang, “ANFIS: adaptive-network-based fuzzy inference system,” IEEE Trans. on Systems, Man, and Cybernetics, vol. 23, no. 3, pp. 665–685, 1993.
[18] Roosa, Kimberlyn, et al. "Short-term forecasts of the COVID-19 epidemic in Guangdong and Zhejiang, China: February 13–23, 2020." Journal of clinical medicine 9.2 (2020): 596.
[19] Al-Qaness, Mohammed AA, et al. "Optimization method for forecasting confirmed cases of COVID-19 in China." Journal of Clinical Medicine 9.3 (2020): 674.
[20] Roosa, K., et al. "Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020." Infectious Disease Modelling 5 (2020): 256-263.
[21] Fanelli, Duccio, and Francesco Piazza. "Analysis and forecast of COVID-19 spreading in China, Italy and France." Chaos, Solitons & Fractals 134 (2020): 109761.
[22] Ghanbari, Behzad. "On forecasting the spread of the COVID-19 in Iran: The second wave." Chaos, Solitons & Fractals 140 (2020): 110176.
[23] S. G. Sepanlou, F. Kamangar , H. Poustchi, R. Malekzadeh, Reducing the Burden of Chronic Diseases: A Neglected Agenda in Iranian Health Care System, Requiring A Plan for Action, Archives of Iranian Medicine, 13(4) (2010) 340-350
[24] Kapoor, Amol, et al. "Examining COVID-19 Forecasting using Spatio-Temporal Graph Neural Networks." arXiv preprint arXiv:2007.03113 (2020).
[25] Zeroual, Abdelhafid, et al. "Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study." Chaos, Solitons & Fractals 140 (2020): 110121.
[26] Mousavi, Seyed Muhammad Hossein. "A New Way to Age Estimation for RGB-D Images, based on a New Face Detection and Extraction Method for Depth Images." International Journal of Image, Graphics and Signal Processing 10.11 (2018): 10.
[27] https://coronavirus.jhu.edu/map.html
[28] https://www.ecdc.europa.eu/en/covid-19/data
[29] https://github.com/CSSEGISandData/COVID-19