ISSN 2278-3091
Volume 9, No.1.4, 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse8591.42020.pdf
https://doi.org/10.30534/ijatcse/2020/8591.42020

The Influences of Global Geographical Climate towards COVID-19 Spread and Death

Shafaf Ibrahim[1], Saadi Ahmad Kamaruddin[2], Nurbaity Sabri[1], KhyrinaAirinFariza Abu Samah[1], MohdRahmatMohd Noordin[1], AnisAmilah Shari[1]
[1]Faculty of Computer and Mathematical Sciences, UniversitiTeknologi MARA, Jasin, Melaka, Malaysia,
shafaf2429@uitm.edu.my, nurbaity_sabri@uitm.edu.my, khyrina783@uitm.edu.my,
mrahmat.noordin@uitm.edu.my, anisamilah@uitm.edu.my
[2]School of Quantitative Sciences, College of Arts & Sciences,Universiti Utara Malaysia,06010 UUM Sintok,
Kedah DarulAman,Malaysia, saadiahmadkamaruddin@gmail.com

ABSTRACT
As the world's coronavirus disease 2019 (COVID-19) case total and death toll continue to climb, an increasing data collection and analysis are providing insights into the pandemic. Although outbreaks continue to develop rapidly, and researchers' understanding of the virus is increasing, a consensus is emerging on certain main aspects of the spread, symptoms, and deadliness of the virus. Enormous global data distribution on COVID-19 is made available online with a combination of global climate data, which creates an opening for further analysis to be conducted. To date, the global climate change has been studied widely, particularly regarding its influences on the distribution of species. This reflects the need for an analysis that is best suited to big data analysis which offers high performance and efficiency in understanding this pandemic issue. The state-of-art in data mining and statistics areas show that the adaptation of these methods could be the most suitable candidate for this purpose. We, therefore, proposed to investigate the influences of the global geographical climate towards the COVID-19 spread and death using a technique of Artificial Neural Network (ANN). It is believed that the proposed study could introduce a new suggestive strategy in improving the precaution measures, enhancing the new normal living activities, and to increase the performance scalability of big data processing comprehensively.

Key words: COVID-19, Global Geographical Climate, Artificial Neural Network (ANN)

1. INTRODUCTION

As of 12 March 2020, coronavirus disease 2019 (COVID-19) was confirmed in 125,048 people worldwide, with a mortality rate of around 3 to 7 percent [1]. By way of the world's COVID-19 case total and death toll continue to climb,an **i**ncreasing data collection and analysis are providing insights into the pandemic.Althoughoutbreaks continue to develop rapidly, and researchers' understanding of the virus is increasing, consensus is emerging on certain main aspects of the spread, symptoms, and deadliness of the virus. What are the aspects that influence the virus's spread? What are the best probabilities and what are their relations associated to the virus' spread?

Geographical distribution refers to the way something is spread out over a geographic area. The distribution principle can be extended to nearly everything on Earth, from animal and plant species, to disease infections, weather patterns, and man-made structures. The scientific community has been looking at the potential impacts of climate change caused by greenhouse gases on biodiversity [2]. Global climate change has been extensively investigated in this context, especially with regard to its effect on the current and potential future distribution of species [3]. Moreover, there is a significant interest in the impacts of global climate change on other biodiversity patterns, particularly in terms of functional diversity and phylogenetic[4]. Several recent research, both locally [5] and globally [6], have attempted to recognize and map future changes in spatial climate variability and to understand the consequences to the conservation planning in a biogeographic context, whereas this aspect appears to have implications for conservation projects due to the direct impact on species survivals [7].

Providentially, enormous global data distribution on COVID-19 is made available online with combination of global climate data, which creates an opening for further analysis to be conducted. This reflects the need for an analysis that is best suited to big data analysis which offers high performance and efficiency in understanding this pandemic issue. The state-of-art in data mining and statistics areas show that the adaptation of these methods could be the most suitable candidate for this purpose. Statistics and data mining techniques are exciting ways to derive information from the data [8]. The use of statistics and data mining has recently been pursued in several research fields [9], [10]. Data Mining is a process that extracts potential, effective, and understandable data from massive data in accordance with established business objectives [11]. At a shallow level, ituses query, recovery and reporting features of the current database management system in conjunction with multi-dimensionalanalysis and statistical analysis methods,

Table 1: Descriptive Statistics

		cases	deaths	HIGH TEMPERATURE	LOW TEMPERATURE	AVERAGE TEMPERATURE	population	TOTAL CASES	TOTAL DEATH	PERCENT CASES POPULATION	PERCENT DEATH POPULATION
N	Valid	41	41	41	41	41	41	41	41	41	41
	Missing	0	0	0	0	0	0	0	0	0	0
Mean		303.8537	8.5366	81.7756	63.9220	72.8756	111331860.0732	11519.7805	363.8293	.0094	.0002
Median		32.0000	.0000	88.7000	69.1000	79.3000	23816775.0000	6991.0000	165.0000	.0000	.0000
Mode		.00	.00	66.20[a]	77.00	83.60[a]	437479.00[a]	4651.00[a]	129.00[a]	.00	.00
Std. Deviation		579.74393	21.18620	13.83014	14.11557	13.39953	303784756.30899	30617.90898	1238.19098	.05739	.00151
Variance		336103.028	448.855	191.273	199.249	179.547	92285178165714976.000	937456350.326	1533116.895	.003	.000
Skewness		3.010	2.989	-.785	-.836	-.751	4.090	6.376	6.388	6.376	6.366
Std. Error of Skewness		.369	.369	.369	.369	.369	.369	.369	.369	.369	.369
Kurtosis		10.588	8.308	-.341	-.134	-.433	16.218	40.762	40.869	40.757	40.666
Std. Error of Kurtosis		.724	.724	.724	.724	.724	.724	.724	.724	.724	.724
Range		2936.00	89.00	52.70	56.10	52.00	1438886297.00	199544.00	8032.00	.37	.01
Minimum		.00	.00	49.00	24.30	36.90	437479.00	2986.00	61.00	.00	.00
Maximum		2936.00	89.00	101.70	80.40	88.90	1439323776.00	202530.00	8093.00	.37	.01
Sum		12458.00	350.00	3352.80	2620.80	2987.90	4564606263.00	472311.00	14917.00	.39	.01
Percentiles	25	1.0000	.0000	72.1500	51.1000	60.6500	5478484.0000	5745.0000	135.5000	.0000	.0000
	50	32.0000	.0000	88.7000	69.1000	79.3000	23816775.0000	6991.0000	165.0000	.0000	.0000
	75	280.0000	4.5000	92.8500	76.0000	84.1000	76896463.5000	8344.5000	187.0000	.0000	.0000

and performs online analytical processing (OLAP) to obtain statistical analysis of the data for decision making [12], [13].

The Artificial Neural Network (ANN) is a network which attempts to mimic human brain neuronal functionality.In the neural network context, a neuron can be understood as a transmitter that interacts with a specific output when stimulated with a specific input or set of inputs [14]. The power, flexibility and ease of use make it as the preferred tool for many predictive data mining applications. Predictive neural networks are particularly useful in applications with complex underlying mechanism, such as predicting control for robot algorithm [15], forecasting of nonlinear time series [16], forecasting solar [17], predicting the total annual crude oil export [18], and many more. ANN is widely used in predictive applications[19], for example, the multilayer perceptron (MLP) and radial basis function (RBF) networks, are controlled in a way that the sense of the model-predicted results can be compared to known target variable values.

We, therefore, proposed to investigate the influences of the global geographical climate towards the COVID-19 spread and death using Artificial Neural Network (ANN). The employment of Artificial Neural Network (ANN) is expected to contribute in understanding the influences of global geographical climate as well as finding an effective suggestive strategy in handling the COVID-19 spread and death in future. It is believed that the proposed study could introduce a new suggestive strategy in improving the precaution measures, enhancing the new normal living activities, and to increase the performance scalability of big data processing comprehensively.

2. DATA BACKGROUND

A COVID-19 dataset which includes the number of cases and death were collected from the European Centre for Disease Prevention and Control(ECDPC), global geographical climate data were taken from the Weather Forecast, and population data is obtained from the Current World Population. The dataset covered the 41 countries in Asia, however, due to incomplete data distribution, three countries were excluded which are Palestine, Tajikistan and Yemen. The descriptive statistics data can be seen in Table 1.

3. RESEARCH METHODS

In this research, the multilayer perceptron model was selected. The ANN was performed using SPSS 23. The similar approach can be seen in [20] – [23]. Two-layer neural network is adapted, with hyperbolic tangent transfer function in the first layer andpurelin transfer function at the second layer. Trainscg is the training function used in this research, with mean squared error (MSE) equals to 0.0 as the criterion function. The data are partitioned into three different sets which are training (70%),validation (30%) and testing (30%) sets. A nonlinear model that contains more than one predictor variableneural network is developed.

The theoretical framework comprises of two variables which are independent and dependent variablesas demonstrated in Table 2. Whereas, Figure 1 illustrates the theoretical framework of this research in orderto achieve the main objective.

Table 2: Type of Variables

Variable	Description	Notation	Type
Dependent	Total Death	TOTALDEATH	Continuous
Independent	Daily Cases	cases	Continuous
	Daily Death	death	Continuous
	High Temperature	HIGH TEMPERATURE	Continuous
	Low Temperature	LOW TEMPERATURE	Continuous
	Population	population	Continuous
	% Cases over Population	PERCENT CASES POPULATION	Continuous
	%Death over Population	PERCENT DEATH POPULATION	Continuous
	Average Temperature	AVERAGE TEMPERATURE	Continuous
	Total Cases	TOTALCASES	Continuous

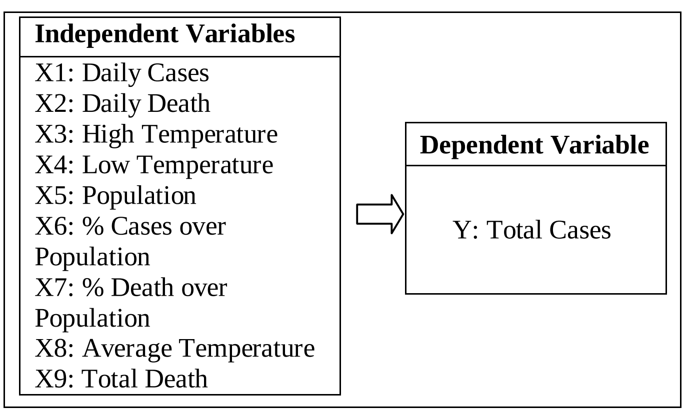

Figure 1: Theoretical Framework

The neural network model with nine predictor variablesin represented in (1).

$$Y = purelin\left[LW^{2,1}\left[\tanh\left\{\begin{array}{l}(IW_1)^{1,1} * X1 + \\ (IW_2)^{1,2} * X2 + \\ (IW_3)^{1,3} * X3 + \\ (IW_4)^{1,4} * X4 + \\ (IW_5)^{1,5} * X5 + \\ (IW_6)^{1,6} * X6 + \\ (IW_7)^{1,7} * X7 + \\ (IW_8)^{1,8} * X8 + \\ (IW_9)^{1,9} * X9\end{array}\right\}\right]\right] \quad (1)$$

In this research, the general model of the COVID-19 spread and death is represented in (2). The final model will include only the significant predictors to describe the total mark.

$$TotalDeath = purelin\left[LW^{2,1}\left[\tanh\left\{\begin{array}{l}(IW_1)^{1,1} * cases + \\ (IW_2)^{1,2} * death + \\ (IW_3)^{1,3} * HIGH \\ TEMPERATURE + \\ (IW_4)^{1,4} * LOW \\ TEMPERATURE + \\ (IW_5)^{1,5} * \\ POPULATION + \\ (IW_6)^{1,6} * \\ PERCENT \\ CASES \\ POPULATION + \\ (IW_7)^{1,7} * \\ PERCENT \\ DEATH \\ POPULATION + \\ (IW_8)^{1,8} * \\ AVERAGE \\ TEMPERATURE + \\ (IW_9)^{1,9} * \\ TOTAL \\ CASES\end{array}\right\}\right]\right] \quad (2)$$

4. RESULTS

The case processing summary is presented in Table 3. Based on Table 3, in the preprocessing part, the data were divided into two sets; training and testing. The training set consist of 75.6 % (31/41) of the overall data, while testing sets comprises of 24.39% (10/41) of the overall data, N=41. There were no excluded values recorded.

Table 3: Case Processing Summary

		N	Percent
Sample	Training	31	75.6%
	Testing	10	24.4%
Valid		41	100.0%
Excluded		0	
Total		41	

Next, Table 4 tabulates the overall network information.

Table 4: Network Information

Input Layer	**Covariates**	1	Cases
		2	Death
		3	High Temperature
		4	Low Temperature
		5	Population
		6	% Cases over Population
		7	% Death over Population
		8	Average Temperature
		9	Total Cases
	Number of Units[a]		9
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		1
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Total Cases
	Number of Units		1
	Rescaling Method for Scale Dependents		Standardized
	Activation Function		Identity
	Error Function		Sum of Squares

a. Excluding the bias unit

The covariates to the network were Cases, Deaths, High Temperature, Low Temperature, Population, Percentage of Cases over Population, and Percentage of Death over Population, Average Temperature, and Total Cases. These nine covariates were theinputs nodes in the input layer of the network. This network consists of only one hidden layer, with one single node.The activation function from input layer to hidden layer was hyperbolic tangent. The target of the network isCOVID-19 spread and death, where the activation function from hidden layer to output layer was identity (purelin). Thedefault error function in backpropagation neural network was based on sum of squares (SSE). To simplify, the configurations of this network was 9-1-1. The network architecture can be referred in Figure 2.

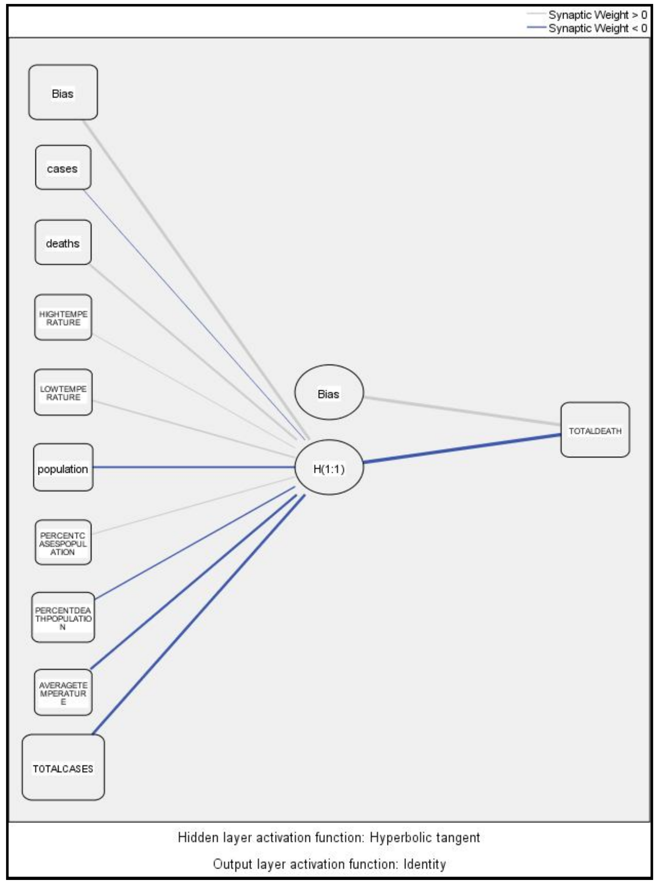

Figure 2: Network Architecture

Subsequently, Table 5 depicts the model summary of both training and testing sets. For this particular network, the sum of squares error fortraining set was 0.045, with relative error equals to 0.003. On the other hand, the sum ofsquares error for testing sets was 0.036, with relative error equals to 9.527. It can be said that in any network, testing set shouldbe the reference. The relative error was 9.527 percent which was quite low. Therefore, it is firmly believed that the network performance is in decent structure.

Table 5: Model Summary

Training	Sum of Squares Error	.045
	Relative Error	.003
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00.01
Testing	Sum of Squares Error	.036
	Relative Error	9.527

Dependent Variable: TOTALDEATH

a. Error computations are based on the testing sample.

Figure 3 and Figure 4 portray the standardized residuals and the distribution of standardized residualsrespectively. It is found that the top five most essential factors which global geographical climate towards COVID-19 spread and death are the total case (100%), percentage of death over population (28.2%), population (16.1%), low temperature (10.1%) and high temperature (7%) as demonstrated in Table 6, and the related figure is shown in Figure 5.

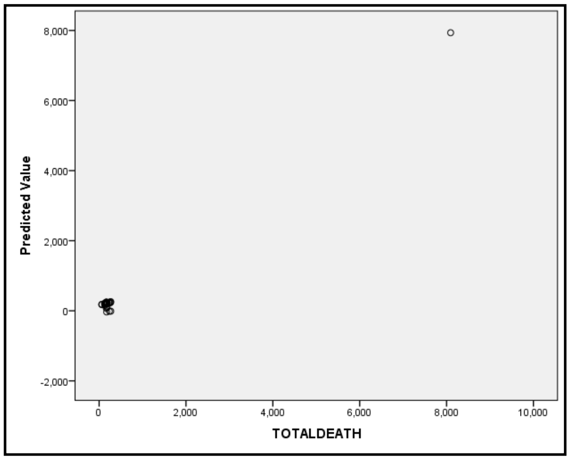

Figure 3: Standadized Residuals

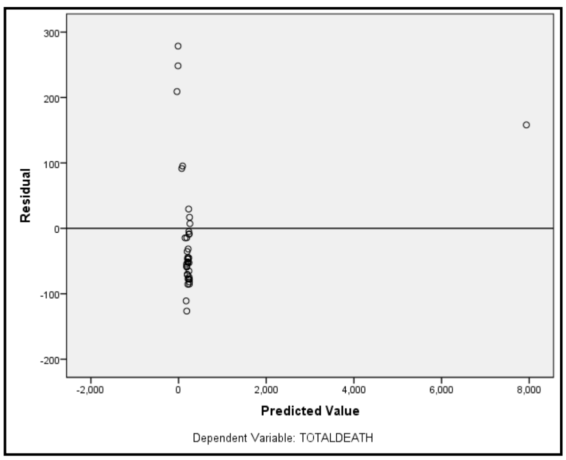

Figure 4: Distribution of Standardized Residuals

Table 6: Independent Variable Importance

	Importance	Normalized Importance
Cases	0.006	1.0%
Deaths	0.021	3.8%
High Temperature	0.039	7.0%
Low Temperature	0.056	10.1%
Population	0.090	16.1%
% Cases over Population	0.019	3.5%
% Death over Population	0.157	28.2%
Average Temperature	0.054	9.8%
Total Cases	0.558	100.0%

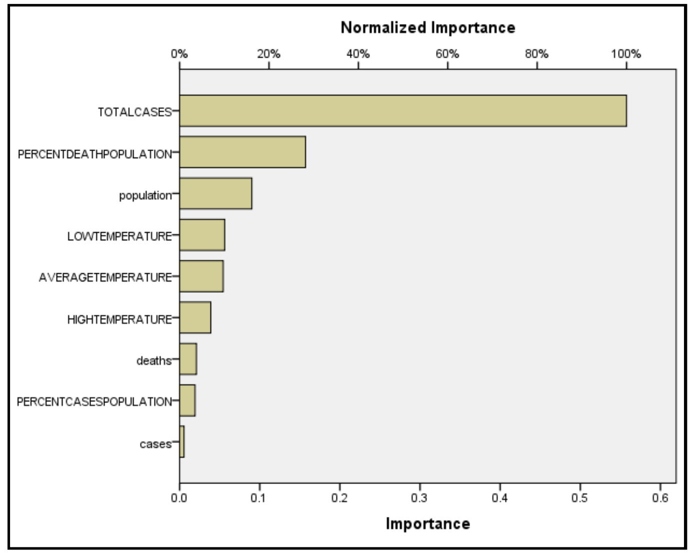

Figure 5: Normalized Importance

5. CONCLUSION

As a conclusion, the study's main objective to investigate the influences of the global geographical climate towards the COVID-19 spread and death using Artificial Neural Network (ANN) has been successfully achieved. The configuration of thenetwork adapted was 9-1-1, with hyperbolic tangent and purelin activation functions in hidden layer and outputlayer respectively. It is expected that optimal results can be achieved as the network performance is in decent structure. In the near future, several ANN models will becompared in order to find the optimal model for the dataset used in this research in orderto avoid any over and under fittingproblem.On the different note, this study is just a preliminary study phase. The dataset and context should be appliedin investigating the influences of global geographical climate towards the COVID-19 spread and deathat different world continental in which the impacts of the global geographical climate could be seen comprehensively.

ACKNOWLEDGMENT

The research was supported by UniversitiTeknologi MARA Cawangan Melaka through the Teja Internal Grant 2020 (GDT2020-19).

REFERENCES

1. M. Puja, F. M. Daniel, B. Michael, S. Emilie, T. Rachel, and J. M. Jessica. **COVID-19: consider cytokine storm syndromes and immunosuppression**, *The Lancet*, vol. 395, issue 10229, pp. 1033-1034, 2020.
2. J. E. Thornes. ***Climate Change 2001: Impacts, Adaptation, and Vulnerability***, IPCC (Intergovernmental Panel Climate Change), Contribution of Working Group II to the Intergovernmental Panel on Climate Change Third Assessment Report, Cambridge: Cambridge University Press, 2002.

3. W. Thuiller.**Biodiversity - Climate change and the ecologist.***Nature*, Vol. 448, pp. 550-552, 2007. https://doi.org/10.1038/448550a
4. T.Wilfried, L. Sébastien, R. Cristina, B. Isabelle, L. Bruno, andA. Miguel. **Consequences of climate change on the tree of life in Europe**. *Nature*, Vol. 470, pp. 531-534, 2011.
5. C. Parmesan. **Ecological and evolutionary responses to recent climate change**.*Annual Review of Ecology, Evolution and Systematics*, Vol.37, pp. 637-690, 2006.
6. J. T. Kerr, H. M. Kharouba, andD. J. Currie. **The macroecological contribution to global change solutions**. *Science*, Vol.316, pp. 1581-1584, 2007.
7. D. D. Ackerly, S. R. Loarie, W. K. Cornwell, S. B. Weiss, H. Hamilton, R. Branciforte,andN. J. B. Kraft.**The geography of climate change: implications for conservation biogeography**. *Diversity and Distributions*, Vol. 16, pp. 476-487, 2010.
8. T.Mauritsius, A. S. Braza,andFransisca. **Bank Marketing Data Mining using CRISP-DM Approach**, *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no.5, pp. 2322-2329, September 2019. https://doi.org/10.30534/ijatcse/2019/71852019
9. R. M. Mohammadand A. Ali. **How statistics and text mining can be applied to literary studies?,***Digital Scholarship in the Humanities*, vol. 34, issue 3, pp. 536–541, September 2019.
10. B. K. Bhardwaj. **A Critically Review of Data Mining Segment: A New Perspective**, *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no.6, pp. 2984-2987, November 2019. https://doi.org/10.30534/ijatcse/2019/50862019
11. H. Shi, L. Zhang, L. Zhengand G. Liu. **Research on Data Mining in Medical Data Visualization**, in *Proceedings of the 2019 8th International Conference on Networks, Communication and Computing (ICNCC 2019)*, New York, 2019, pp. 79–82.
12. L. Asma, E. Brahim, E. Malika, and S. Abdallah. **Data mining techniques application for prediction in OLAP cube**, *International Journal of Electrical and Computer Engineering*,vol. 9, issue 3, pp. 2094-2102, 2019.
13. Q. Lu, Z. Lyu, Q. Xiang, Y. Zhou, and J. Bao.**Research on data mining service and its application case in complex industrial process**, in *2017 13th IEEE Conference on Automation Science and Engineering (CASE)*, China, 2017, pp. 1124-1129.
14. A. Ansari, M. Heras, J. Nones, M. Mohammadpoor, andF. Torabi.**Predicting the performance of steam assisted gravity drainage (SAGD) method utilizing artificial neural network (ANN),***Petroleum*, pp. 1-7, 2019.
15. Y. Wang. **Robot algorithm based on neural network and intelligent predictive control**.*Journal of Ambient Intelligence and Humanized Computing*, 2019.
16. A. Tealab, H.Hefny, andA. Badr. **Forecasting of nonlinear time series using ANN**, *Future Computing and Informatics Journal*, vol. 2, issue 1, pp.39-47, 2017.
17. B. Sivaneasan, C.Y. Yu, andK.P. Goh, **Solar Forecasting using ANN with Fuzzy Logic Pre-processing**, *Energy Procedia*, Vol. 143, pp. 727-732, 2017.
18. T. Alam.**Crude Oil Export: A Comparative Analysis using Artificial Neural Network and ARIMA**, *International Journal of Advanced Trends in Computer Science and Engineering*, volume 8, no.5, pp. 2546-2550, September 2019. https://doi.org/10.30534/ijatcse/2019/102852019
19. O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, andG. Li. **An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction**, *Expert Systems with Applications*, Vol. 68, pp. 163-172, 2017.
20. M. Yin, J. Ma, J. Xu, L. Li, G. Chen, Z. Sun, andW. Mo. **Use of artificial neural networks to identify the predictive factors of extracorporeal shock wave therapy treating patients with chronic plantar fasciitis**, *Scientificreports*, vol. 9, no. 1, pp. 1-8, 2019.
21. K. Zhang, Y. Hu, andY. Wu. **Classification and Prediction on Rural Property Mortgage Data with Three Data Mining Methods**, *Journal of Software Engineering and Applications*, vol. 11, no. 07, pp. 348-361, 2018.
22. R. J. LaFaro, S.Pothula, K. P. Kubal, M. E. Inchiosa, V. M. Pothula, S. C. Yuan, D. A. Maerz, L. Montes, S. M. Oleszkiewicz, A. Yusupov, andR. Perline. **Neural network prediction of ICU length of stay following cardiac surgery based on pre-incision variables**, in *PLoS One*, vol. 10, no.12, pp. 1-19, 2015.
23. A. S. Mahmoud, R. S. Farag, M. M. Elshfai, L. A. Mohamed, andS. M. Ragheb. **Nano Zero-Valent Aluminum (nZVAl) Preparation, Characterization, and Application for the Removal of Soluble Organic Matter with Artificial Intelligence**, *Isotherm Study, and Kinetic Analysis. Air, Soil and Water Research*, Vol. 12, pp. 1-13, 2019. https://doi.org/10.1177/1178622119878707