Research Article

International Journal of Advanced Technology and Engineering Exploration, Vol 8(74)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2020.S1762140

Students' learning habit factors during COVID-19 pandemic using multilayer perceptron (MLP)

Nur Nabilah Abu Mangshor[1*], Shafaf Ibrahim[1], Nurbaity Sabri[1] and Saadi Ahmad Kamaruddin[2]
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Melaka (Jasin Campus), 77300 Merlimau, Melaka, Malaysia[1]
School of Qualitative Sciences, College of Arts & Sciences, Universiti Utara Malaysia, 06010, UUM Sintok, Kedah, Malaysia[2]

Received: 16-November-2020; Revised: 25-January-2021; Accepted: 26-Janjuary-2021

Abstract
Rapid dissemination of coronavirus disease 2019 (COVID-19) across the globe has necessitated the introduction of social distance interventions to slow the spread of the disease. Online learning has become essential, considering the implications of this virus to be spread among the students during physical classes. Hence, educational institutions have shifted the traditional physical classes to online classes. Due to this implementation worldwide, a study on student learning habits is crucial to analyse students learning habits as it is one of the main factors that affecting students' performance in learning. Fifteen independent variables as inputs to one of the well-known Artificial Neural Network algorithms, Multilayer Perceptron (ANN-MLP) has been used to investigate the student's learning habit factors during the COVID-19 pandemic. Through analysing original survey data from 420 secondary students (Grade 6-12) in Hanoi shows that the ANN-MLP model is stable for both ANN-MLP optimization algorithms which are for Adjusted Normalized, to be concise. The hours spend for self-learning before COVID-19 is observed to be the most influential factors of student's learning habit during COVID-19 pandemic. Moreover, the promising Sum of Squares Error (SSE) and Relative Error (RE) values obtained signify that the ANN-MLP model is appropriate in identifying the student's learning habit factors during COVID-19 pandemic.

Keywords
COVID-19, Learning habits factor, Artificial neural network (ANN), Multilayer perceptron (MLP).

1.Introduction
The emerge of the COVID-19 pandemic around the world gave a high impact especially in education sectors where the conventional approach not able to be practice due to this pandemic [1]. There is an urge for the implementation of online learning to cope with the current situation and to avoid the disease from spreading. Online learning is a learning alternative where it provides a synchronous or asynchronous learning experience through the internet with the leveraging of time and place flexibility between both students and instructors [2]. However, the sudden changes in educational content to the digital world may affect student learning habits [3]. In this paper, a factor that contributes towards student learning habits during the COVID-19 pandemic will be analysed.

Students learning habits are very important because they influence the performance and quality of learning [4]. Learning habits are defined as a method of gathering, organizing, the activities related to psychological, cognitive, and affective fields of interaction with learning environments [5]. During the COVID-19 pandemic, online learning influences the student learning habits including personal interests for learning, problem-solving, competencies, and project working skills [6]. The algorithm inspired by the diverse structure of the human brain; the Artificial Neural Network (ANN) has been used in numerous complex problems [7]. It can process a large volume of data, learn from training data, and excellent generalization capability [8]. Due to its capability of self-learning and self-adapting, more research on this algorithm has been successfully implemented to solve real-world problems. It also able to perform more efficient and accurate numerous classification [9−11], prediction [12, 13] and many

*Author for correspondence

more compared to other classification techniques [14].

The Multi-Layer Perceptron (MLP) is one of the well-known Artificial Neural Networks (ANNs) where data with expected outputs are applied to this network also known as supervised training [15]. MLP outperforms the Linear Regression in the prediction of student performance in the final examination [16]. MLP is also recommended in predicting a student achievement based on motivation, learning, and emotional intelligence where it gives a high number of performance accuracy compared to the Decision Tree algorithm [17]. MLP outperforms Random Forest (RF) in student performance prediction where it able to produce more fine and meticulous results compared to RF. Meanwhile, a comparison between MLP, Support Vector Machine (SVM), K-Nearest Neighbour (KNN), and Discriminant (DISC) to predict student awareness in ICT and Mobile technology shows a remarkable result. The MLP and SVM also outperform KNN and DISC with no significant differences in performance accuracy. It is recommended to be used for real-time prediction [18].

Due to the positive correlation between both student learning habits and performance, investigation on the factor that contributes to the student learning habits during the COVID-19 pandemic is necessary. Hence, due to the promising results and advantages of MLP in self-learning towards a real-world problem, this research employs MLP in determining the contributory factors of Student's Learning Habit during COVID-19 pandemic. In this context, the objectives of this paper are as follow:

1. To present a MLP network model in investigating the contributory factor of students' learning habits during the COVID-19 pandemic
2. To evaluate the performance of the proposed MLP network model using the standard Sum of Squares Error (SSE) and Relative Error (RE).

Following, this research paper contains the Introduction which is presented in Section 1 and followed by Data Background in Section 2. Next, Section 3 and Section 4 discuss the Methods and Results, respectively. Subsequent, Section 5 presented the Discussion and Section 6 deliberated the Conclusion and Future Work.

2.Data background

A dataset on Vietnamese students' learning habits during the COVID-19 pandemic [3] was retrieved from the ResearchGate open source. This dataset is established from the questionnaires distributed in the period of 7th February 2020 to 28th February 2020. It is composed of three major groups of questions which are (1) student's demographic, (2) student's learning habits, and (3) student's perceptions of self-learning during the school closure. This dataset is composed of 460 responses from the secondary student in grades 6-12. However, only 420 data are valid after the process of data cleaning. There were fifteen covariates used in the network which *are fa_job, exam, Self_evaluation, English, Lh_before_Cov, nec_prog, nec_habit, nec_parent, eff_moti, eff_con eff_supp, eff_env, eff_obj, eff_resource,* and *eff_friend.*
Table 1 tabulates the description of the covariates used in this study.

Table 1 Covariates description

Input layer	Covariates	Descriptions
1	*fa_job*	What is your father's job?
2	*exam*	What subject group are you intend to take for university entrance exam?
3	*Self_evaluation*	How do you evaluate your performance, regarding your selected subject group?
4	*English*	How do you evaluate your English capability?
5	*Lh_before_Cov*	Before COVID-19, how many hours do you spend per day for self-learning?
6	*nec_prog*	I can assure my learning progress
7	*nec_habit*	I can maintain my learning habit
8	*nec_parent*	My parents show me it is necessary
9	*eff_moti*	I have motivation for self-learning
10	*eff_con*	I have proper concentration skill
11	*eff_supp*	I have support from my family
12	*eff_env*	I have an effective learning environment
13	*eff_obj*	I can define my daily learning objectives
14	*eff_resource*	I have various learning resources
15	*eff_friend*	I communicate and collaborate with my friends about learning

3.Methods

This section discusses the methodology used in conducting this study. This study aims to evaluate the performance of the MLP network is investigating the contributory factors of students learning habits during the COVID-19 pandemic. *Figure 1* depicts the proposed methodology used in this study.

From *Figure 1*, the Vietnamese students' learning habit dataset is used as the input to the model. Next, data normalization and cleaning are performed. Consequently, the process of initialize and train the MLP model is conducted. The model is constructed using SPSS 13. In general, the structure of ANN is composed of an input layer, hidden layers, and the output layer [19]. Next, *Figure 2* represents MLP network architecture used in this study. From the architecture, xn represents the first layers of the network which is known as the input layer. The prediction result produces by the output layer (third layer) is obtained from a calculation made by the hidden layer (second layer). These three layers consist of a few nodes where each of these nodes connect with one another on each layer. Each of the nodes on each layer is known as a neuron except for the input layer where the nodes x1, x2, x3 and xn are addressed as the input features to the feed to the network.

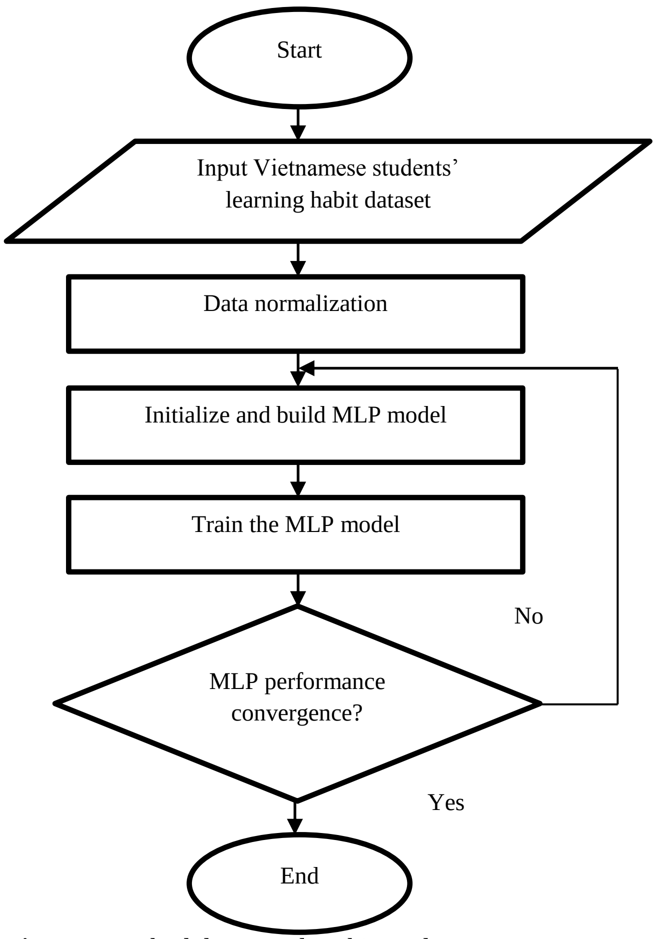

Figure 1 Methodology used in this study

Each neuron consists of three main components which are input signals represented with x1 until xn, multiply by the weight represented by w1 until wn. The input sum is a calculation of weighted input plus a bias (b) in a linear equation as shown in equation (1). Next, the activation functions where determined whether the neuron can be activated or not as output y. The additional of bias to the weighted of sum produces the activation function. Following shows the linear equation for artificial neuron:

$$y = f\left(b + \sum_{i=1}^{n} w_i\, x_i\right) \qquad (1)$$

In general, the data source factor chosen will affects the number of input layer nodes in the network. On the other hand, the number of hidden neurons in the network is determined based on the specified training dataset used. Next, hidden layers are used for the calculation purpose by applying the dedicated function. Each node in the input layer must connected to all the nodes located in the hidden layer and the nodes at the hidden layer must be connected to all the nodes at the output layer [20].

In this study, the fifteen covariates acted as the input for the MLP network. The activation function used in the network is *Softmax* and there is one dependent variable which is *Lh_in_Cov*. To simply the architecture, it can be addressed as 15-7-1. The MLP network architecture as portrayed in *Figure 3.*

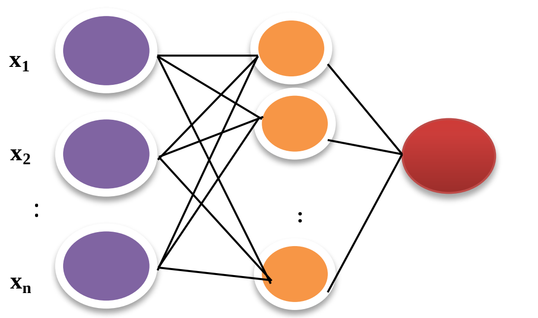

Figure 2 MLP network architecture

Based on the illustration of the MLP network architecture in *Figure 3*, this network is composed of 15 input and one hidden layer with seven nodes. Softmax function is used as the hidden layers' activation function. Hence, the target output of this network is the learning hour in covid (Lh_in_Cov). Identity (purelin) function is used as the activation function from the hidden layer to the output layer and the default error function applied is the SSE. SSE is the sum of differences between each observation group means, and it is used to indicate the variation within a cluster. Hence, it can be concluded that the smaller the SSE, the lesser the variation between the cluster.

4.Results

This section discusses the results obtained from the experiment conducted. The case processing summary for MLP network is portrayed in *Table 2*. Based on *Table 2*, the dataset used in this study is divided into two groups which are training and testing. The training set for MLP is 66.7% (280/420) and the testing set is 33.3% (140/420). The overall data from the dataset is N = 420 and there is no excluded data reported. Next, *Figure 4* and *Figure 5* portray the standardized residuals and the distribution of standardized residuals respectively. It is concluded that the five most influential factors of students' learning habit during COVID-19 are *Lh_before_Cov* (100), *eff_obj* (54.9), *eff_con* (53.1), *English* (41.4), and *eff_moti* (34.4) as shown in *Table 3.*

Following, *Figure 6* portrays the normalized importance independence variables in a different representation. All the 15 covariates are presented according to it normalize importance.

As mentioned previously, the performance evaluation was conducted on the developed MLP network model using the SSE and RE. *Table 4* and *Table 5* demonstrate the SSE and RE for the MLP network model, respectively. Further elaboration on both *Table 4* and *Table 5* are deliberated in Section 5 Discussion accordingly.

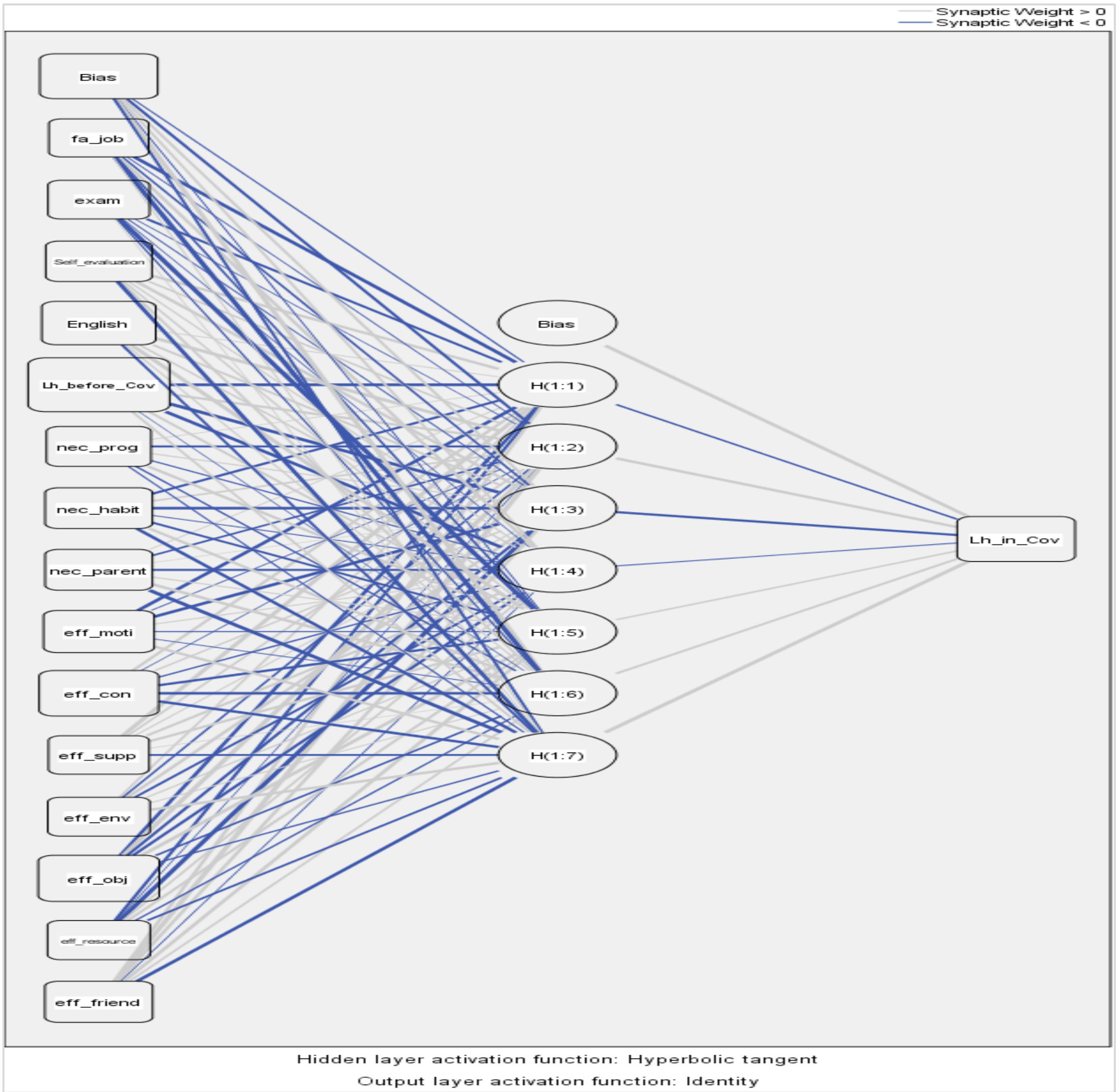

Figure 3 MLP network architecture

Table 2 Case processing summary for RBF model

		N	Percent
Sample	Training	280	66.7%
	Testing	140	33.3%
Valid		420	100.0%
Excluded		0	
Total		420	

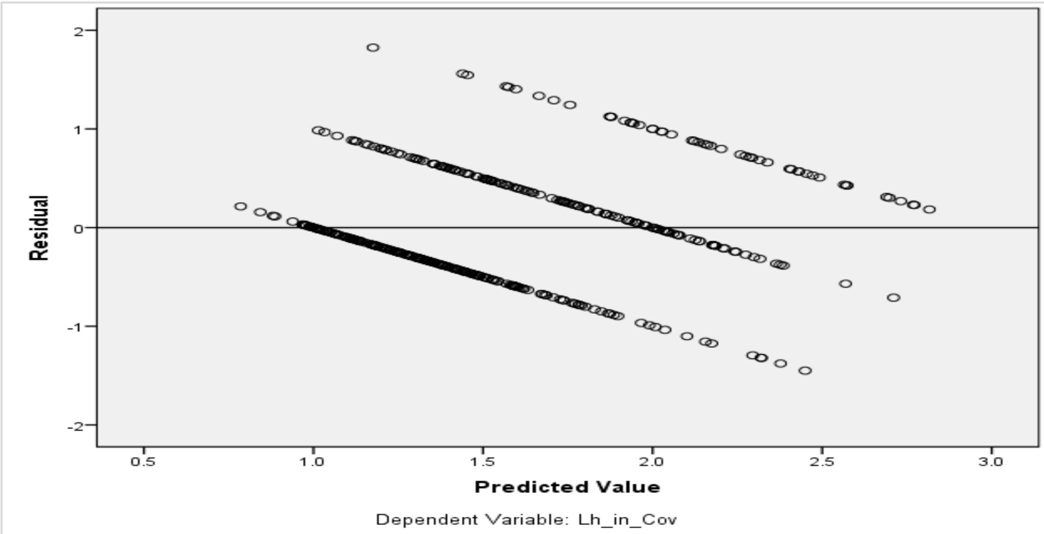

Figure 4 Standardized residuals

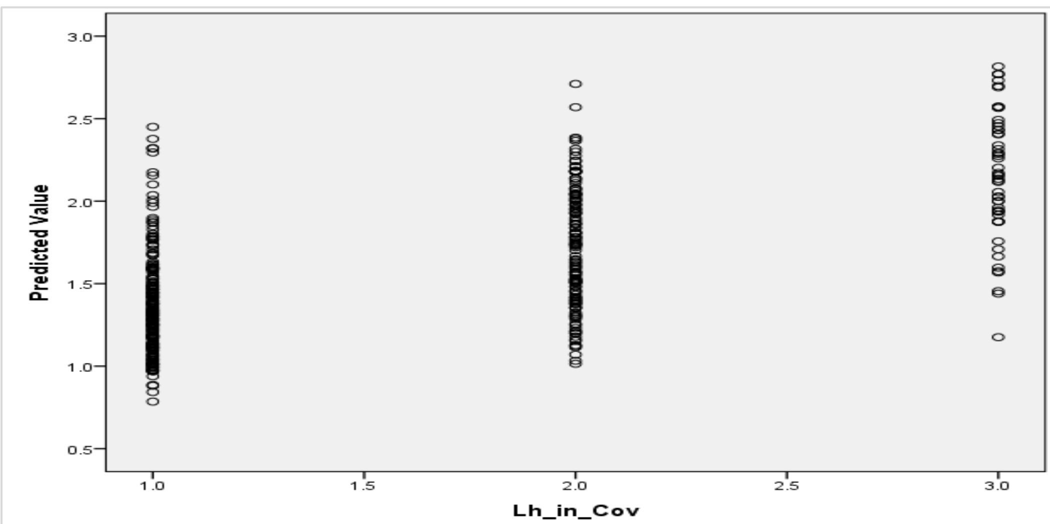

Figure 5 Distribution of standardized residuals

Table 3 Independent variables' importance

Variables	**Importance**	**Normalized importance**
What is your father's job?	.032	19.8%
What subject groups do you intend to take for the university entrance exam?	.063	38.9%
How do you evaluate your performance regarding your selected subject group?	.065	40.5%
How do you evaluate your English capability?	**.065**	**40.7%**
Before COVID-19, how many hours did you spend per day for self-learning?	**.161**	**100.0%**
I can assure my learning progress	.064	39.8%
I can maintain my learning habit	**.072**	**44.7%**
My parents show me it is necessary	.044	27.7%
I am motivated for self-learning	**.088**	**54.9%**
I have proper concentration skills	**.081**	**50.5%**
I have support from my family	.045	28.1%
I have an effective learning environment	.058	36.4%

Variables	Importance	Normalized importance
I can define my daily learning objectives	.064	39.5%
I have various learning resources	.059	36.8%
I communicate and collaborate with my friends about learning	.038	23.4%

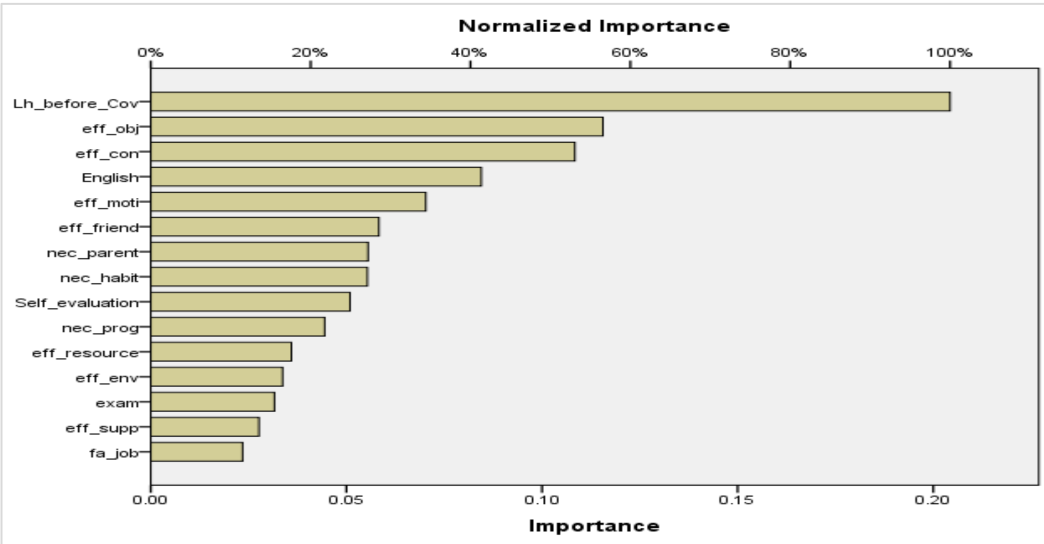

Figure 6 Normalized importance

Table 4 SSE of ANN-MLP model

Rescaling of covariates	Optimization algorithm	
	Scaled conjugate gradient	Gradient descent
Standardized	102.813	97.762
Normalized	95.731	96.218
Adjusted Normalized	85.564	88.562
None	97.848	102.669

Table 5 RE of ANN MLP model

Rescaling of covariates	Optimization algorithm	
	Scaled conjugate gradient	Gradient descent
Standardized	0.679	0.650
Normalized	0.674	0.666
Adjusted Normalized	0.613	0.602
None	0.677	0.696

5.Discussion

Referring to the SSE results tabulated in *Table 4,* it can be observed that the Adjusted Normalized produced the lowest SSE rate for both Scaled Conjugate Gradient and Gradient Descent which are 85.564 and 88.562 respectively. The None rescaling on the other hand returned the highest SSE in Gradient Descent which is 102.669. Also, the Standardized by Scaled Conjugate Gradient is monitored to be the least efficient by returning the highest SSE rate at 102.813.

Next, similar for RE, the Adjusted Normalized recurrently yield the lowest rate for both Scaled Conjugate Gradient and Gradient Descent which are 0.613 and 0.602 correspondingly in *Table 5.* Therefore, it can be concluded that the MLP model is stable for both MLP optimization algorithms which is for Adjusted Normalized, to be particular. These promising SSE and RE values obtained signify that the MLP model is appropriate in identifying the Students' Learning Habit Factors during COVID-19 pandemic.

6.Conclusion and future work

This study investigates the contributory factor on students learning habits during the COVID-19 pandemic using the Artificial Neural Network of Multilayer Perceptron (ANN-MLP) model. The proposed model was implemented on a Vietnamese dataset. Fifteen covariates were used in this research which is based on a 15-7-1 structure. The major contribution to students' learning habits is English capability, motivation for self-learning, concentration skill, daily learning objective, and the most major contribution is hours spend for self-learning before COVID-19 with 100% normalized importance. From the result obtained, it shows that ANN-MLP is an appropriate model to identify the contributory factors for the students' learning habits with a promising rate of SSE and RE. By understanding the contribution factors, it could improve teaching and learning processes, and overcome the unprecedented challenges in teaching and learning during COVID-19 pandemic.

In a nutshell, this study implements the original MLP network. The hybridization of the existing MLP model with any AI techniques in future is believed to offers expansion of knowledge in this technique. In addition, the use of other statistical approaches despite SSE and RE will provides deeper insights in understanding the model performance.

Acknowledgment

All authors would like to express appreciation to Universiti Teknologi MARA Cawangan Melaka for providing financial support for this research under the TEJA Internal Grant Scheme 2020 (GDT2020-36).

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1] Adnan M, Anwar K. Online learning amid the COVID-19 pandemic: students' perspectives. Online Submission. 2020; 2(1):45-51.
[2] Singh V, Thurman A. How many ways can we define online learning? a systematic literature review of definitions of online learning (1988-2018). American Journal of Distance Education. 2019; 33(4):289-306.
[3] Trung T, Hoang AD, Nguyen TT, Dinh VH, Nguyen YC, Pham HH. Dataset of Vietnamese student's learning habits during COVID-19. Data in Brief. 2020:1-7.
[4] Urh M, Jereb E. Learning habits in higher education. Procedia-Social and Behavioral Sciences. 2014; 116:350-5.
[5] Reid JM. Learning styles in the ESL/EFL classroom. Heinle & Heinle Publishers, International Thomson Publishing Book Distribution Center, 7625 Empire Drive, Florence, KY 41042.; 1995.
[6] Bhebhe S, Maphosa C. Examining the learning habits of distance education learners in one Southern African University. Asian Journal of Distance Education. 2020; 15(1): 257-68.
[7] Hoffmann LF, Bizarria FC, Bizarria JW. Detection of liner surface defects in solid rocket motors using multilayer perceptron neural networks. Polymer Testing. 2020.
[8] Venu K, Palanisamy N, Krishnakumar B, Sasipriyaa N. Disease identification in plant leaf using deep convolutional neural networks. In handbook of research on applications and implementations of machine learning techniques 2020 (pp. 46-62). IGI Global.
[9] Saritas MM, Yasar A. Performance analysis of ANN and naive bayes classification algorithm for data classification. International Journal of Intelligent Systems and Applications in Engineering. 2019; 7(2):88-91.
[10] Feng X, Ma G, Su SF, Huang C, Boswell MK, Xue P. A multi-layer perceptron approach for accelerated wave forecasting in lake Michigan. Ocean Engineering. 2020.
[11] Cui X, Wang Q, Zhao Y, Qiao X, Teng G. Laser-induced breakdown spectroscopy (LIBS) for classification of wood species integrated with artificial neural network (ANN). Applied Physics B. 2019; 125:1-12.
[12] Nasser IM, Al-Shawwa MO, Abu-Naser SS. Developing artificial neural network for predicting mobile phone price range. 2019; 3(2):1-6.
[13] Feng R, Gao H, Luo K, Fan JR. Analysis and accurate prediction of ambient PM2. 5 in china using multi-layer perceptron. Atmospheric Environment. 2020.
[14] Al-Mubayyed OM, Abu-Nasser BS, Abu-Naser SS. Predicting overall car performance using artificial neural network. 2019; 3(1):1-5.
[15] Olmedo MT, Paegelow M, Mas JF, Escobar F. Geomatic approaches for modeling land change scenarios. Springer International Publishing; 2018.
[16] Widyahastuti F, Tjhin VU. Predicting students performance in final examination using linear regression and multilayer perceptron. In international conference on human system interactions (HSI) 2017 (pp. 188-92). IEEE.
[17] Fahri MU, Isa SM. Data mining to prediction student achievement based on motivation, learning and emotional intelligence in MAN 1 Ketapang. International Journal of Modern Education and Computer Science. 2018; 10(6):53-60.
[18] Verma C, Stoffová V, Illés Z. Prediction of students' awareness level towards ICT and mobile technology in Indian and Hungarian University for the real-time: preliminary results. Heliyon. 2019; 5(6):1-9.
[19] Zhou T, Jiang Z, Liu X, Tan K. Research on the long-term and short-term forecasts of navigable river's water-level fluctuation based on the adaptive multilayer perceptron. Journal of Hydrology. 2020.

[20] Pham BT, Nguyen MD, Bui KT, Prakash I, Chapi K, Bui DT. A novel artificial intelligence approach based on multi-layer perceptron neural network and Biogeography-based optimization for predicting coefficient of consolidation of soil. Catena. 2019; 173:302-11.

Nur Nabilah Abu Mangshor is a lecturer in Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Melaka, Kampus Jasin. She holds a Diploma, Bachelor's Degree, and Masters in Computer Science. Her research interests are Artificial Intelligence and Image Processing.
Email: nurnabilah@uitm.edu.my

Dr. Shafaf Ibrahim is a Senior Lecturer in Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (Malacca) Jasin Campus. Have a qualification of Diploma, Bachelor's Degree, Masters and PhD in Computer Science. Her research interests are Artificial Intelligence, Evolutionary Algorithms, Image Processing, Machine Learning, and Optimization.
Email: shafaf2429@uitm.edu.my

Nurbaity Sabri is a lecturer in Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Melaka, Kampus Jasin. She holds a Master in Computer Science from Universiti Teknologi Malaysia (UTM). She is a programming lecturer and currently teaching Computer Programming in C++ and Java programming.
Email: nurbaity_sabri@uitm.edu.my

Dr. Saadi bin Ahmad Kamaruddin is a senior lecturer at School of Quantitative Sciences (SQS), University Utara Malaysia, Sintok campus. He received both of his Master's Degree in Applied Statistics, and PhD in Information Technology and Quantitative Sciences from UiTM. He was a bachelor's graduate in Mathematics with Hons from IIUM. His specialties are Artificial Neural Network, Metaheuristic Algorithm, and Data Analytics.
Email: s.ahmad.kamaruddin@uum.edu.my