ISSN: 0067-2904

Improving Measurement of Effectiveness of Blended Learning in Iraqi Education Using SVM

Reem Razzaq Abdul Hussein [1]*, Huda Dheyauldeen Najeeb [2]

[1] *Department of Business Information Technology, College of Business Informatics, University of Information Technology and Communications, Baghdad, Iraq*

[2] *Department of Public Relations, College of Media, Al Iraqia, University, Baghdad, Iraq*

Received: 10/12/2021 Accepted: 26/1/2022 Published: 30/9/2022

Abstract

E-learning has recently become of great importance, especially after the emergence of the Corona pandemic, but e-learning has many disadvantages. In order to preserve education, some universities have resorted to using blended learning. Currently, the Ministry of Higher Education and Scientific Research in Iraq has adopted e-learning in universities and schools, especially in scientific disciplines that need laboratories and a spatial presence. In this work, we collected a dataset based on 27 features and presented a model utilizing a support vector machine with regression that was enhanced with the KNN method, which identifies factors that have a substantial influence on the model for the type of education, whether blended or traditional.
Furthermore, the dataset used was primarily focused on three key factors: personal information, the impact of e-Learning platforms, and the influence of the Corona virus. The attributes that were measured revealed that social status, computer skills, and the basic platform gave the user enough tools to continue the learning process. The size of the classrooms and laboratories that meet the health safety conditions is the most significant. The goal of our work is to discover a model that predicts how blended learning will be used during and after the coronavirus pandemic and to produce a model with minimal errors.

Keywords: Machine learning, SVM with regression, KNN, Feature selection, Blending learning.

تحسين قياس فاعلية التعلم المدمج في التعليم العراقي باستخدام SVM

ريم رزاق عبد الحسين [1]*, هدى ضياءالدين نجيب [2]

[1] قسم تكنولوجيا معلومات الاعمال، كلية معلوماتية الأعمال, جامعة تكنولوجيا المعلومات والاتصالات، بغداد، العراق

[2] كلية الاعلام، قسم العلاقات العامة, الجامعة العراقية، بغداد، العراق

*Email: huda_iraq81@yahoo.com

الخلاصة

في الآونة الأخيرة ، أصبح للتعلم الإلكتروني أهمية كبيرة ، خاصة بعد ظهور جائحة كورونا ، على الرغم من ان للتعلم الإلكتروني عيوب كثيرة. لذلك, من أجل الحفاظ على التعليم ، لجأت بعض الجامعات إلى استعمال التعلم المدمج. ففي الوقت الحاضر ، اعتمدت وزارة التعليم العالي والبحث العلمي في العراق التعلم الإلكتروني في الجامعات والمدارس ، وخاصة التخصصات العلمية التي تحتاج إلى مختبرات ووجود مكاني.

في هذا العمل ، جمعنا مجموعة بيانات تستند إلى 27 ميزة وقدمنا نموذجًا يستعمل آلة متجه الدعم مع الانحدار الذي تم تحسينه باستعمال طريقة K-NN. والتي تحدد العوامل التي لها تأثير كبير على النموذج لنوع التعليم ، سواء كانت ممزوجة أو تقليدية.

علاوة على ذلك ، ركزت مجموعة البيانات المستعملة بشكل أساسي على ثلاثة عوامل رئيسية : المعلومات الشخصية ، وتأثير منصات التعلم الإلكتروني ، وتأثير فيروس كورونا. السمات التي تم قياسها تبين أن الوضع الاجتماعي ومهارات الحاسوب والنظام الأساسي يوفر للمستخدم أدوات كافية لمواصلة عملية التعلم ، وحجم الفصول الدراسية والمختبرات التي تلبي شروط السلامة الصحية هي اهم السمات. الهدف من عملنا هو اكتشاف نموذج يتنبأ بكيفية استعمال التعلم المدمج أثناء وبعد جائحة فيروس كورونا ، وإنتاج نموذج بأقل قدر من الاخطاء

1. Introduction

The COVID-19 pandemic is rapidly spreading across the globe and shows no signs of slowing down. There are various infection modes[1]. Furthermore, the epidemic has impacted many aspects of the economy and politics, as well as people's social lives all around the world [2][3]. The internet has generated new habits and routines of life as a result of the social separation and isolation required to prevent the epidemic from spreading [4][5].

Education is one of the fields that has been heavily affected because it is essential for many people in many different countries [6]. Because students' capacity to attend school and participate in group activities is limited, exchange activities, research, and most learning is conducted through blended learning or e-learning approaches [4][7].

E-learning has seen substantial expansion in recent years and is now attracting a growing number of students[8]. E-learning courses make use of the internet's capabilities to enhance and improve existing educational methodologies[9]. As well as supporting individual learners in acquiring knowledge in a flexible manner outside of the confines of a classroom, this innovative method of instruction also benefits educational institutions, which can now serve more students while lowering their costs[8][10].

This paper focuses on e-learning in Iraqi universities by building a model that predicts how blended learning will be used during and after the coronavirus pandemic by using machine learning techniques. The paper is organized as follows. In section 2, literature discussion about work-related e-learning is included. In section 3, the support vector machine is discussed. In section 4, the design of the proposed method is explained. Section 5 discusses dataset collection. The results and implementation, discussion, and conclusion of the paper are shown in sections 6 and 7, respectively.

2. Related work

The COVID-19 pandemic is affecting many aspects of society, including university training programs around the world [11]. As a result, many universities now employ online learning as a viable option. There have been many studies dealing with e-learning[12], including:

Sucitra [13] presented a model to determine the pass rate of e-learning students by using support vector machine (SVM). This model consists of several stages: 1) Data Processing to remove duplicates and delete data of the same value or redundant. 2) Data validation to identify and remove odd data (noise /outliers), incomplete data (missing values), and inconsistent data by using particle swarm optimization (PSO). 3) Support Vector Machine

(SVM) for classification and decision. The dataset was taken from the Moodle e-learning platform database, which is used for the teaching and learning process at Bina Sarana Informatika University in Indonesia. Data obtained from data collection is 400 records, which have six attributes: gender, number of exercises, number of logins, number of document access, number of forums, and number of messages.

It was found that the number of exercises was the attribute that most influenced student graduation results, with a probability value of 0.898. The results of this model prove that the support vector machine (SVM) has a very good generalization ability to solve problems even with a limited sample, with an accuracy value of 81.02% and an AUC (Area Under Curve) value of 0.810.

Aditya et al. [14] presented an approach for classifying e-learning systems by using Least Square–Support Vector Machine (LS-SVM) and neural networks (NN). The e-learning database of the National Center for Biotechnical Information (NCBI) was used for classification, which includes 240 samples and eight types of attributes: Student prior knowledge (SPK), Study level (SL), Motivation (M), Grades from previous semester (GP), Cognitive style (C), Learning style (L), Personality (P), and Anxiety (A). The experiment results show that the highest accuracy is 73.12%, obtained from employing the LS-SVM method.

Dang et al. [15] proposed a model for choosing one of the learning methods (e-learning, traditional or blended learning) by using several methods of machine learning: naïve bayes (NB), support vector machine (SVM), K-nearest neighbor (KNN), random forest (RF), and multilayer perceptron (MP). In this study, 679 data samples were obtained from 303 students at Vietnam's Academy of Journalism and Communication (AJC). The model was based on four influential factors (courses; infrastructure; teacher; student), which included 16 features. According to the results of the experiment: 1) the largest impact on the system is the infrastructure factor, then teachers, courses, and students, respectively. 2) The random forest (RF) method achieved the highest accuracy (81.52%).

This model was compared with Alberto et al. [16] model, which was based on five events and included 39 features. The result was that the proposed model [15] gave 3.32% higher accuracy than model [16].

Elin et al. [17] presented comparison results of assessment student learning results with e-learning lectures by first combining particle swarm optimization (PSO) with support vector machines (SVM) and then with neural networks. Nine types of attributes are used for evaluating gender, quizzes, log totals, collecting assignments, messages, uploading total tasks, collecting assignments, discussion, and chat. The finding of the research was that the approach gives high accuracy (97.35%) when PSO is combined with a neural network, which is superior to combining PSO with a support vector machine by approximately 8.88%.

Kuo et al. [18] employed three machine learning algorithms: support vector machine (SVM), deep belief network (DBN), and decision tree for classifying e-learning students' courses. E-learning customer marketing datasets contain 198 records used for classification, of which 148 for training and 50 for testing. Each record has six attributes: home postal code, private and public university, birth month, gender, and decision variable class of study. The finding of the research was that business courses were almost always chosen by female students, while male students almost always chose engineering courses, except for those who were born after September and lived in Taipei or those who were born after March and lived south of Changhua. The inclination of students to study e-learning courses is unaffected by whether they attend a private or public university. The results show that the support vector machine (SVM) algorithm achieves a higher testing accuracy (81.8%) than other commonly used algorithms.

3. Support vector machine (SVM)

The Support Vector Machine (SVM) is a collection of connected learning algorithms for recognizing patterns and analyzing data, as well as regression analysis and classification [19]. SVM provides a noise-resistant solution by taking a collection of input data and predicting for each of the inputs, which are divided into two classes and classified using the best hyperplane value [20][21], which is shown in Figure 1. The SVM's fundamental concept is that to distinguish data from two different classes depending on the features of the data, it should find the best hyperplane such that the distance between the two classes, i.e., the margin, is maximized. A simple linear maximum margin classifier is used to create the basic SVM classifier in the following manner: [22][23]

$$D = (A_1, B_1), (A_2, B_2),\ldots, (A_M, B_M) \quad (1)$$

$$D= (A_n, B_n), \forall n = 1, 2,\ldots, M \quad (2)$$

For each n = 1,2,.., M, A_n is input vector and $B_n \in \{+1,-1\}$ is the associated class label.

Important objects that should be tracked are indicated by positive values, while everything else that should not be tracked is indicated by negative values. [20][24].

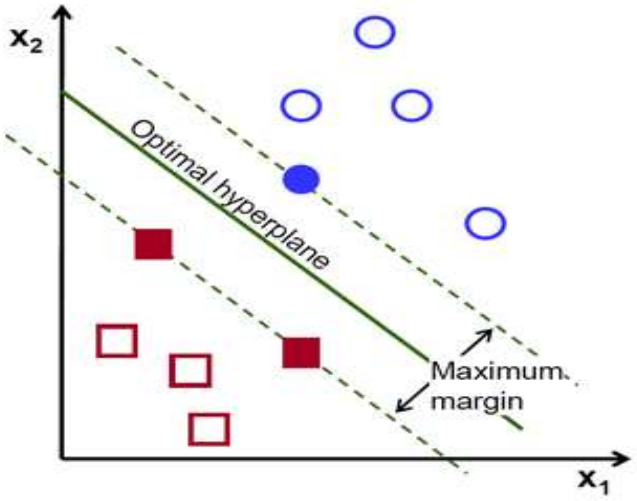

Figure 1- SVM classifier [25]

4. Proposed Method Design

In this circumstance, we need to measure whether blended education will be feasible following the Corona pandemic. The dataset collected is collected from the various Iraqi universities' lectures. The proposed work applied to 738 records. The proposed algorithm can be explained as a flowchart, as shown in Figure 2.

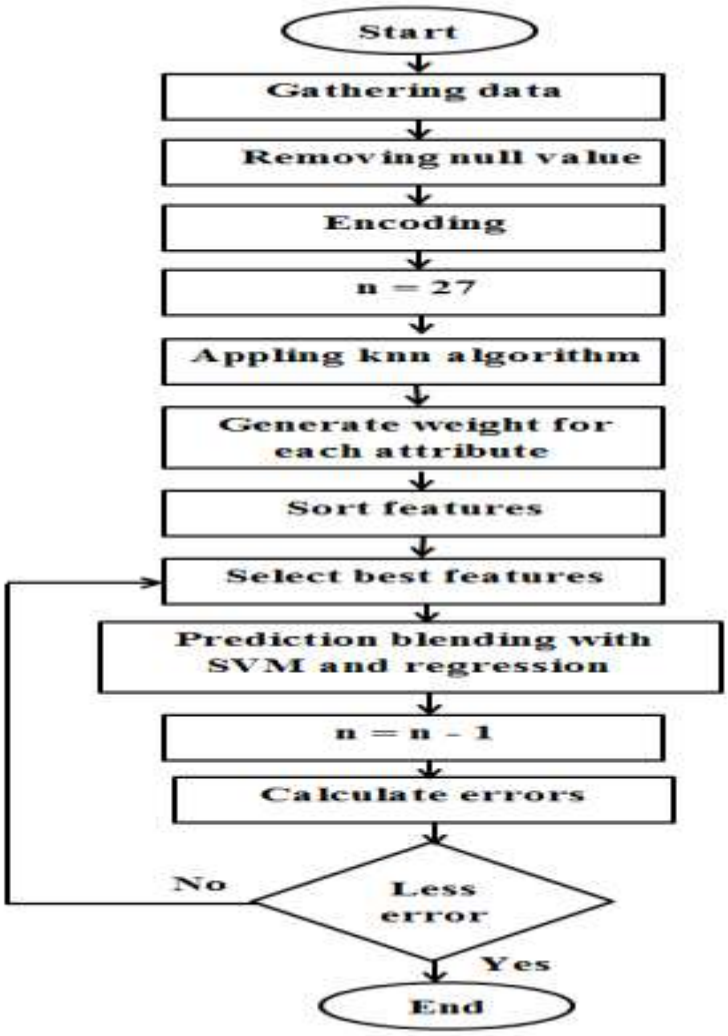

Figure 2- Proposed Method Design

The major steps of the suggested algorithm's work:

1. The input of the model is collected data, which contains 515 records.
2. Preprocessor removes null values, incompleteness, and inconsistencies in the dataset. Thus, the restrictions were reduced in our dataset from 515 to 497 samples. This phase is important to reduce the errors caused by the null value and to increase the accuracy of the final model .
3. A model is constructed by applying a support vector machine regression to predict how blended learning will be used during and after the coronavirus pandemic. The accuracy of the SVM classifier was based on RMSE.
4. Evaluation and validation of results: the application of the proposed work is carried out several times to determine the level of model accuracy. Each time a set of features is selected, the error rate is calculated until the best features with the lowest error rate are reached.
5. The output of the model: producing a model with minimal errors which is used for predicting the blended

5. Dataset collection and description

Data was collected according to the questionnaire of the various Iraqi universities' professors. The questionnaire was created using Google Form, where each questionnaire consisted of 28 questions, and it was distributed to 515 professors (samples) at various universities. The

questionnaire was described below in (Table 1). The dataset was cleaned by removing null values and noise. The preprocessing has been applied by using the KNN method for feature selection.

Table 1- Iraqi Universities' Professors Dataset

Feature Category	Question	Feature	Description	Feature Number
Personal Information	Q1	Governorate	Text	**1**
	Q2	Gender	Binary (Female, Male)	**2**
	Q3	Age	Numeric (1: 22 - 35, 2:36-50, 3:51-65 year)	**3**
	Q4	Educational level	Nominal (Bc, M.Bc, PhD)	**4**
	Q5	Personal income	Nominal (moderate, Good, Vgood)	**5**
	Q6	Social status	Nominal (single, married, other)	**6**
	Q7	Specialization	Binary (scientific , Humanitarian)	**7**
Platforms Impact Factor	Q8	Are you using the zoom platform, FCC, or Google Meet?	Ordinal (FCC, Google, MEET, I do not use any platform)	**8**
	Q9	Do you use Moodle platform, Google class Room, Edmodo	Ordinal (Moodle, Google class Room, Edmodo, I do not use any platform)	**9**
	Q10	Is the class divided into small groups?	Binary (Yes, No)	**10**
	Q11	Do you create video recordings for students?	Binary (Yes, No)	**11**
	Q12	Do you communicate with students on an ongoing basis?	Nominal (Yes, No, Sometimes)	**12**
	Q13	Do you find any difficulties in learning modern programs?	Binary (Yes, No)	**13**
	Q14	After the Corona pandemic, you prefer traditional or Blended Learning	Binary (Traditional, Blended Learning)	**28**
	Q15	Do you have computer skills?	Binary (Yes, No)	**14**
	Q16	Have you approved the change of course according to the pandemic circumstances?	Nominal (Yes, No, Almost)	**15**
	Q17	Is the default environment appropriate?	Binary (Yes, No)	**16**
	Q18	Are you interested in learning new platforms?	Binary (Yes, No)	**17**
	Q19	Do you want to develop the skills of using the platform of your organization?	Binary (Yes, No)	**18**
	Q20	Does the platform provide the user enough tools to continue the learning process?	Binary (Yes, No)	**19**
	Q21	Do you have children in the evaluation stage and do they suffer from e-learning?	Binary (Yes, No)	**20**
The Influence of the Coronavirus Factor	Q22	Does your institution have the conditions for the safety and prevention of the Corona pandemic?	Binary (Yes, No)	**21**
	Q23	Does the size of the classrooms and laboratories meet the health safety conditions?	Binary (Yes, No)	**22**

Q24	Are students obligated to attend electronically electronic lectures?	Nominal (Yes, No, Sometimes)	**23**
Q25	Are students obligated to attend with regards to the lectures given in the college?	Nominal (Yes, No, Sometimes)	**24**
Q26	Have you or any of your family members had a previous infection with Coronavirus?	Binary (Yes, No)	**25**
Q27	Do you have a fear of the Corona pandemic?	Binary (Yes, No)	**26**
Q28	Do you have chronic diseases?	Binary (Yes, No)	**27**

:

6. Results and discussion

The major contributions that were provided in the proposed work will be considered in this section. Our study relied on the KNN algorithm. Use the neighbourhood weighting approach to determine the most significant feature that influences the model's accuracy. Each feature carries its own weight, as seen in Table (2). Each feature has a different weight, and the feature with the most weight in the proposed model is the most efficient.

Table 2-The Weight of each Feature

Weight	Feature Number
2.226415	27
2.18388	14
2.106498	3
1.919558	6
1.897383	25
1.637761	9
1.300337	4
1.188229	16
1.186431	22
1.04826	8
0.979232	26
0.824313	19
0.623451	24
0.585326	10
0.580087	15
0.523586	11
0.000473	2
0.000128	1
7.41E-05	12
4.03E-05	20
3.64E-05	7
2.89E-05	17
1.91E-05	23
1.89E-05	21
1.77E-05	18

1.24E-05	13
5.57E-06	5

The feature selection threshold is chosen through trial and error. First, all features are applied to the model, and then the feature is eliminated from m=26 to m=13, as shown in Table 3. From the table, it was noted that the minimum error appears in step(3) when choosing 24 features from 28 features, where the RMSE becomes 0.00013651.

We noticed that the SVM with regression is a powerful method for measuring effective blending learning, but it works more robustly with the KNN method, as the error is minimized from (0.00014) to (0.00013).

Table 3-Results of hybridization SVM and KNN Algorithms

Number of Features	RMSE of SVM
26	0.00014325
25	0.00014486
24	0.00013651
23	0.00014229
22	0.00014217
21	0.00014306
20	0.00013972
19	0.00013797
.18	0.00013653
17	0.00014139
16	0.00013868
15	0.00014033
14	0.0001415
13	0.00014463

The proposed work was compared with research[15], which used the SVM classification technique on different datasets. Research [15] aims to select the method of learning (e-learning, traditional or blended learning) based on students at Vietnam's Academy of Journalism and Communication. The result was that our model's prediction accuracy was higher than that of the previous study [15], where the comparison of the results is summarized in the following table.

Table 4- Comparison Proposed work with Research [15]

Criterion	Proposed work	**Research** [15]
Number of records	515 records	679 records
Algorithm for feature selection	Used KNN algorithm for selection	Used K-mean algorithm for selection
Fitting model	One classification method (SVM with regression) was used and was improved which gave strength to the proposed work	Several classification methods (NB, SVM, KNN, RF, and MP) were used and then a method was chosen that gives the best result.
Number of feature and attribute of each factor	Used three factors (personal information, e- Learning platforms impact factor, and influence of the coronavirus factor) with attribute (7,14, and 7) respectively.	Used four factors (student, teacher, infrastructure, and courses) with attribute (6,1, 4, and 5) respectively.
Total of attribute	27 attributes	16 attributes
Dataset	Dataset was collected from the	Dataset was collected from the

	various Iraqi universities' lecturers	Vietnam Academy's students
Performance	good accuracy	accuracy of SVM is less of proposed work
Software used	Using MATLAB version 2020, which gives more flexibility for modification and enhancement	(WEKA) tool version 3.7.8 which is a ready-made tool that is difficult to modify and improve upon

The SVM is discussed in this study. The features that have a significant impact on the model for the type of education, whether blended or traditional, are: social status; those who have computer skills; for whom the platform provides the user with enough tools to continue the learning process; and the size of the classrooms and laboratories that meet the health and safety conditions. While the features of less weight that were excluded from the proposed work are educational level, personal income, the lecture who has children in the assessment stage who suffer from e-learning, and change, of course, according to the pandemic circumstances.

7. Conclusion

Data mining is widely used, especially in the e-learning field. In this paper, the Iraqi dataset was adopted that depends on three main factors, such as: personal information; E-learning platform impact factor; and the influence of the coronavirus factor. The 27 attributes that had been tested were found to be social status, computer skills, and platform, which provide the user with enough tools to continue the learning process. The size of the classrooms and laboratories that meet the health safety conditions is the most important factor that affects the proposed work. The goal of a hybrid model is to determine the impact of inclusion education in the context of coronavirus infection and even beyond. The hybrid model that combines support vector machine regression with the KNN algorithm is based on the important features that contribute to achieving better accuracy, where the error decreased from (0.00014) to (0.00013).

8. Future work

The future selection process can be enhanced by a hybrid with other mining methods.

References

[1] P. Karunanayake and I. D. Hospital, "The COVID-19 Pandemic," vol. 51, no. 1, pp. 1–3, 2020.

[2] I. Chakraborty and P. Maity, "COVID-19 outbreak: Migration, effects on society, global environment and prevention," *Sci. Total Environ.*, vol. 728, p. 138882, 2020, doi: https://doi.org/10.1016/j.scitotenv.2020.138882.

[3] D. Rodriguez-leyva and G. N. Pierce, "The impact of nutrition on the covid-19 pandemic and the impact of the covid-19 pandemic on nutrition," *Nutrients*, vol. 13, no. 6, pp. 1–9, 2021, doi: 10.3390/nu13061752.

[4] C. Rapanta, L. Botturi, P. Goodyear, L. Guàrdia, and M. Koole, "Online University Teaching During and After the Covid-19 Crisis: Refocusing Teacher Presence and Learning Activity," *Postdigital Sci. Educ.*, vol. 2, no. 3, pp. 923–945, 2020, doi: 10.1007/s42438-020-00155-y.

[5] V. A. J. Van Lint and D. Ph, "COVID-19 PANDEMIC ANALYSIS," pp. 1–29, 2021.

[6] R. Razzaq Abdul Hussein, Z. F. Hamza, and B. T. Sabri, "Forecasting the number of COVID-19 infections in Iraq using the ARIMA model," *J. Appl. Sci. Eng.*, vol. 24, no. 5, pp. 729–734, 2021, doi: 10.6180/jase.202110_24(5).0006.

[7] S. M. Leahy, C. Holland, and F. Ward, "The digital frontier: Envisioning future technologies impact on the classroom," *Futures*, vol. 113, no. May, p. 102422, 2019, doi:

10.1016/j.futures.2019.04.009.

[8] I. Lykourentzou, I. Giannoukos, V. Nikolopoulos, G. Mpardis, and V. Loumos, "Dropout prediction in e-learning courses through the combination of machine learning techniques," *Comput. Educ.*, vol. 53, no. 3, pp. 950–965, 2009, doi: 10.1016/j.compedu.2009.05.010.

[9] "E-Learning and its Socioeconomics Avni Singh."

[10] B. Holder, "An investigation of hope, academics, environment, and motivation as predictors of persistence in higher education online programs," *Internet High. Educ.*, vol. 10, no. 4, pp. 245–260, 2007, doi: 10.1016/j.iheduc.2007.08.002.

[11] M. A. Therib, A. F. Al-Baghdadi, and H. A. Marzog, "Medical remotely caring with COVID-19 virus infected people using optimized wireless arm tracing system," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 18, no. 6, pp. 2886–2893, 2020, doi: 10.12928/TELKOMNIKA.v18i6.16331.

[12] N. Hayatin, K. M. Ghufron, and G. W. Wicaksono, "Summarization of COVID-19 news documents deep learning-based using transformer architecture," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 19, no. 3, pp. 754–761, 2021, doi: 10.12928/TELKOMNIKA.v19i3.18356.

[13] S. Antar, B. Vol, and V. I. No, "Penerapan Metode Support Vector Machine (SVM) Guna Menentukan Tingkat Lulus Mahasiswa E-Learning," no. 2, pp. 121–127, 2017.

[14] A. Khamparia and B. Pandey, "SVM and PCA Based Learning Feature Classification Approaches for E-Learning System," *Int. J. Web-Based Learn. Teach. Technol.*, vol. 13, no. 2, pp. 32–45, 2018, doi: 10.4018/IJWLTT.2018040103.

[15] D. N. Lu, H. Q. Le, and T. H. Vu, "The factors affecting acceptance of e-learning: A machine learning algorithm approach," *Educ. Sci.*, vol. 10, no. 10, pp. 1–13, 2020, doi: 10.3390/educsci10100270.

[16] A. Rivas, A. González-Briones, G. Hernández, J. Prieto, and P. Chamoso, "Artificial neural network analysis of the academic performance of students in virtual learning environments," *Neurocomputing*, vol. 423, no. xxxx, pp. 713–720, 2021, doi: 10.1016/j.neucom.2020.02.125.

[17] E. S. Panca Saputra and Indriyanti, "Comparison of Data Mining In E-Learning Learning Based On Log Aktivity On PSO-Based Nural Network Algorithms With PSO-Based SVM," *Indones. J. Artif. Intell. Data Min.*, vol. 3, no. 2, pp. 95–102, 2020.

[18] K. Lin, Y. Lu, and C. Jen, "Analysis of E-learning Customer Data using Data Mining Techniques," no. May, pp. 307–311, 2020.

[19] A. M. S. Rahma, "Cuneiform symbols recognition by support vector machine (SVM)," *J. Al-Qadisiyah Comput. Sci. Math.*, vol. 11, no. 1, 2019, doi: 10.29304/jqcm.2019.11.1.449.

[20] K. R. Reddy, K. H. Priya, and N. Neelima, "Object Detection and Tracking - A Survey," *Proc. - 2015 Int. Conf. Comput. Intell. Commun. Networks, CICN 2015*, no. May, pp. 418–421, 2016, doi: 10.1109/CICN.2015.317.

[21] D. C. Parvathi P1, "An Analysis of Short Text Detection and Classification Algorithms," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 8, no. 4, pp. 176–182, 2020.

[22] A. K. Nandi and H. Ahmed, *Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines*. John Wiley & Sons, 2020.

[23] A. R. Abbas and A. R. Kareem, "Age estimation using support vector machine," *Iraqi J. Sci.*, vol. 59, no. 3, pp. 1746–1756, 2018, doi: 10.24996/IJS.2018.59.3C.19.

[24] A. J. Almahdi, A. J. Yaseen, and A. F. Dakhil, "EEG signals analysis for epileptic seizure detection using DWT method with SVM and KNN classifiers," *Iraqi J. Sci.*, vol. 2021, pp. 54–62, 2021, doi: 10.24996/ijs.2021.SI.2.7.

[25] H. Ahmed and A. K. Nandi, "Compressive sampling and feature ranking framework for bearing fault classification with vibration signals," *IEEE Access*, vol. 6, pp. 44731–44746, 2018.