

Implementation of Efficient Online English Learning System and Student Performance Prediction Using Linear K-Nearest Neighbors (L-Knn) Method

K. Kashinath¹, & Dr. R. L. N. Raju¹

¹Department of English, School of Social Sciences and Languages, Vellore Institute of Technology, Vellore, India

Correspondence: Dr. R.L.N.Raju, Department Of English, School of Social Sciences and Languages, Vellore Institute of Technology, Vellore, India. E-mail: raju.rln@vit.ac.in

Received: February 15, 2022

Accepted: March 25, 2022

Online Published: April 7, 2022

doi:10.5430/wjel.v12n3p235

URL: <https://doi.org/10.5430/wjel.v12n3p235>

Abstract

Technical assistance for the establishment of a distance learning environment for learning English is provided by the advancement of information technology and the educational information process. People are still getting used to online teaching methods, and it is becoming more widely accepted. E-learning and online education have advanced significantly in recent years. The teaching paradigm has moved from traditional classroom learning to dynamic web-based learning. As a result, instead of static information, learners have received dynamic learning material tailored to their abilities, requirements, and preferences. To improve the English learning material efficiency, this paper implements an online English learning system based on efficient learning material selection. The English learning materials are preprocessed using normalization. The dimensionality reduction of the data is done using the Kernel-based-Independent Component Analysis (K-ICA). Data classification is performed using the Hypothetical Naïve Bayes Algorithm (HNBA). The student performance like learning efficiency, interactive accuracy rate, and artistic skills are predicted using the linear k-Nearest Neighbors (L-KNN). The proposed system can be simulated by employing the MATLAB tool and the performance is compared with other conventional methodologies. The findings of this study reveal that the presented online learning method may significantly increase students' oral and written skills.

Keywords: online education, English learning system, Kernel based-Independent component Analysis (K-ICA), Hypothetical Naïve Bayes Algorithm (HNBA), Linear k-Nearest Neighbors (L-KNN)

1. Introduction

The ambition among English learners will always be an interest of researchers as among the most significant study topics regarding English (second language) acquisition (Palvia, S., Aeron, P., Gupta, P., Mahapatra, D., Parida, R., Rosner, R. and Sindhi, S 2018). With the development of the digital age upon this online platform as well as the evolution of distance education items, web-based English programs throughout China, as well as variations on learning, have received much interest among all areas of culture. Nevertheless, study findings on inspiration in conventional language classrooms remain hard to adapt in web-based English classroom students' educational contexts, and the diverse incentives of internet second language learners have deteriorated (Lin, L.F 2018). Traditional static hypothesis struggles to describe & investigate the concept of dynamic growth. As a consequence, the issue of de-motivation between web-based English program students in distant learning was rapidly emerging as a major subject in the area of English language study.

The creation of such a wireless distance educational network built on a Smartphone-based English teaching and learning framework not just to enhance participants' awareness as well as reinforces interactions among both educators and pupils, but that also allows everybody to interact & gain knowledge with one another regardless of time or area (Guo, J., Huang, F., Lou, Y., and Chen, S 2020). The education system in recent times has already had significant ramifications throughout the education sector. Learners' cognitive structure, learning strategies, interpersonal interactions, and self-evaluation are all changing in education. The Smartphone-based wireless remote education setting fulfills the goal of enabling people to study effectively. The phrases "loss" and "lost" were similar among Chinese learners of foreign languages. Their associated study was weaker in extent as well as the frequency

when matched to theirs. This is still in its beginnings in Western nations. The majority of their study was observational. The feeling of irritation throughout the EFL learning classroom setting was comprehensive as a doctorate article concentrating on contributing variables; however, there aren't too many study possibilities that haven't yet included motivational components and learning tools.

Professionals from local and international have conducted extensive research on distance education. Supported by a study & evaluation of existing online or offline education conditions, (McKinley, J 2017) developed an online/offline training method that relies on mobile social networking sites. This approach helps to build the bond connecting instructors and learners. The introduction of innovative online or offline education approaches has also been aided by coordination and teamwork. (Jassim, L.L 2020) They utilize the deployment of the online and offline e-commerce framework to explain universities & colleges' internet application of main education. They utilize WeChat as a framework, incorporate advanced technology into conventional teaching configurations, as well as explore the teaching process of constructing an offline & offline instructional model accompanied by WeChat.

This work implements an online English learning framework related to the effective choice of materials. The materials for the English Learning were standardized. The data are dimensionally reduced employing the K-ICA approach. The HNBA method was utilized to classify the data. The L-KNN estimates the student's achievements like learning efficiency, interaction correctness, and creative abilities. The developed model is simulated with MATLAB as well as its effectiveness compared to existing methods. As an outcome of the research, students' oral as well as reading skills have improved. The additional part of this work is split as follows: topic II-literature survey and problem statement, topic III-proposed work, topic IV-performance analysis, and topic V-conclusion.

2. Literature Survey

Here, we illustrate the existing reports regarding online English learning. (Wang, P. and Qiao, S 2020) This article analyses the English virtual educational environment of landscape architecture courses with maximum security. Video instructional course materials, teacher-student interactivity, English learning in landscape architecture courses, including English teaching materials for landscape architecture majors are supplied online via the environment, the purpose would be to provide students with an effective online learning system. The online English teaching subsystem for landscape architecture majors employs an online assembly & executing application.

(Alqahtani, S.M., Bhaskar, C.V., Vadakalur Elumalai, K. and Abumelha, M 2018) This research examined students' attitudes toward using WhatsApp to learn English. This report's findings suggest that using WhatsApp as a communication tool may result in implicit understanding within and without the classroom. To use WhatsApp for learning effectively, learners must be motivated to discover, understand, and participate in educational activities.

(Bai, H. and Zhang, Q 2020) Online classes are a completely new notion that can lead to intriguing studies. Instructors and learners will benefit from the Innovative Class Schooling System. The laptops, the web, screens, and chalkboards are all used in multi-method classes to help students learn English. From primary school to university level, this educational approach is employed done in a variety of professions and courses. Information about teaching development of smart classrooms IoT applications and 5G networks have the potential to usher in the era of 5G, and IoT analyses it to develop and create different education sectors based on smart gathering ideas. The IoT smart aggregate can suggest new instructional methods based on the English smart classroom.

(Cope, B., Kalantzis, M. and Sears, D 2020) They present a few possible solutions, first theoretically, then operationally, in a summary of the findings of several test deployments that have been reported in considerable detail elsewhere. The main result is that Artificial Intelligence (AI) would never "carry over" the position of educator in the framework of digital estimating practices improved in the last decades because how that serves and what it does are so essentially diverse from human brilliance. Although, within the restraints described in the report, it can change teaching in methods that, possibly confusingly, create more personal, rather than less.

(Xu, Y 2021) They utilize a method to efficiently combine the recovered monomodal characteristics and execute voice assistant, taking into account the significance and compatibility among voice and writing forms. Furthermore, for the industrial Translated version system design, use the edge computing method. Lastly, the Simulation review showed the effectiveness of the industrial Translated version methodology proposed in this research.

(Li, Z 2021) Learning English using the Net is also a very beneficial process of learning English. Learners can learn to read and write for themselves using the web. Learner's attitudes to language models to understand the Technology to search languages and web training to an even good outlook than discovered between many technological advances and alter training techniques allow pupils' attitudes toward language models to understand the Web to learn other languages and web education to a better outlook than noticed between. Learning over the web can be beneficial.

Expanding student networks has a favorable impact on students' possibilities.

(Sun, X 2021) They would use the web and AI innovation to look into how to research a huge (or mega huge), clear and convenient, and high-performing academic instructional data system that can address the needs of educators and pupils no matter what year, month, or day they visit. Carefully optimize and modify several issues faced throughout the system's construction.

(Li, X 2017) They designed and made an intelligent English learning tool using the suppression subtractive hybridization (SSH) architecture. The understanding of intelligent machines in the AI was artistically used to English knowledge point learning and memory procedure. The neural network method was applied in training whenever the human brain's remembering rule was addressed. They evaluate the current software and hardware components till they accomplish the objective.

3. Problem statement

The expansion of digital technologies as well as the teaching-learning processes helps build a remote learning atmosphere for English. Online teaching continues to remain new as well as gaining acceptance. The online system has grown dramatically over the years. Regular classroom education is given way towards flexible web-based education. Rather than semantic information, students nowadays have dynamic content adapted to their talents, needs, and aspirations.

4. Proposed Work

In this phase, we describe the proposed methodology with background performance. This work implements an online English learning framework related to the effective choice of materials. The materials for the English Learning were standardized. The data are dimensionally reduced employing the K-ICA approach. The HNBA method was utilized to classify the data. The L-KNN estimates the student's achievements like learning efficiency, interaction correctness, and creative abilities. Figure 1 depicts the design of the proposed research.

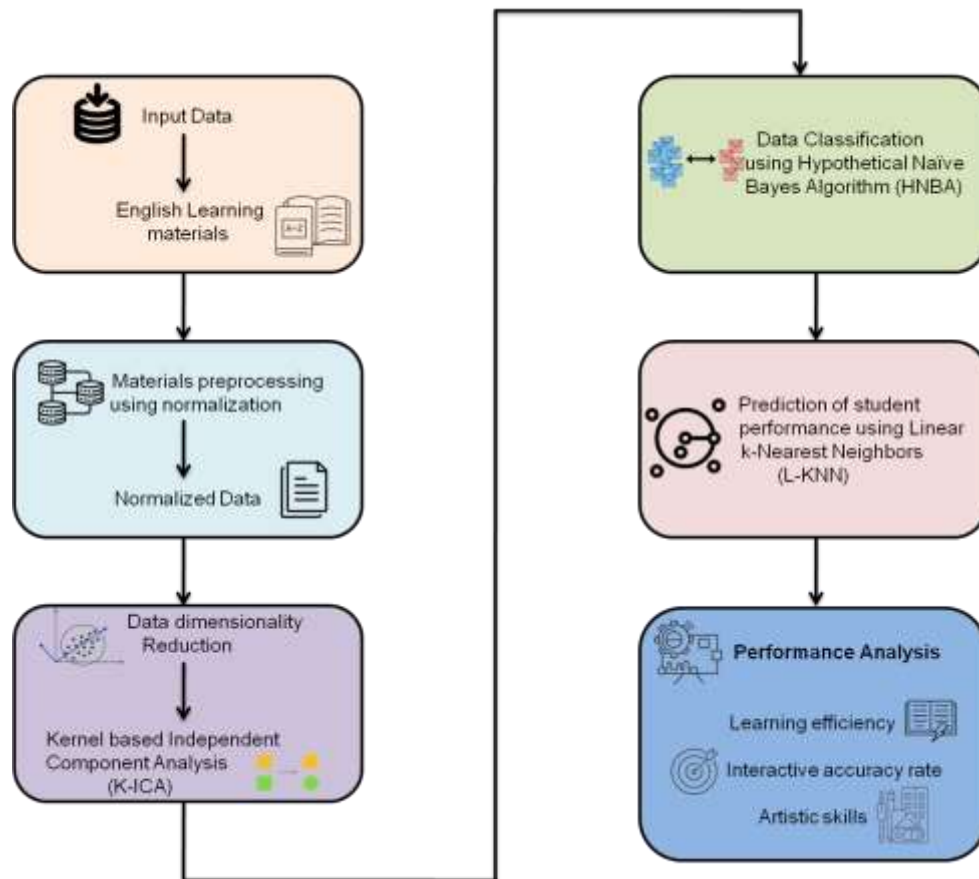


Figure 1. Proposed framework

A. Dataset

This article's corpus information comes out of a university's network education service infrastructure (Zhang, X., Luo, C., He, T., Yang, X., Lu, Z. and Huang, B 2017). There are three sub-forums throughout the online training discussion: unified examination questions, learning support services, and a training forum. The training discussion was separated into seven sub-forums: instructing approach debate, teaching grade books connection, teaching presentations, everyday English interaction, training system assistance program, learning briefing conversation, and also training chapter analysis.

B. Pre-processing using normalization

The learning materials received are unfiltered and will include generated information and inadequate information. It's that has been purified and normalized to delete repeated and redundant distortions, along with the inadequate material. Since the files for the university community are so large, sample compaction techniques must be employed. Because this material dataset has several features, image retrieval methods are needed to sort out the ones which aren't significant. The material dataset may be normalized during the pre-processing stage.

C. Dimensionality reduction using the Kernel-based Independent Component Analysis (K-ICA)

The conversion of such pre-processed high dimensional (HD) picture into a comprehensible depiction with decreased dimension is referred to as dimension mitigation. The dimension of the picture throughout many real-world applications is extremely high, and dealing with such information necessitates reducing it without sacrificing information. The term "higher dimension" relates to the difficulties that come along with multivariate data processing. Although dealing with the kind of large picture necessitates dimension reduction, the purpose of dimension reduction would be to present the picture in a lower-dimensional area while preserving several of the picture's unique features. This dimension mitigation process was carried out by employing the K-ICA approach and this approach converts the HD pre-processed image into a low-dimensional image effectively.

K-ICA is a standard concept to ICA that enhances regular interactions inside a reproducing kernel Hilbert space. Throughout this part, we explain two contrast indexes as well as the accompanying K-ICA procedures.

The initial dimension represents the quantity of observed data (separate two-dimensional pictures belonging to every group). The source image is located in a lower-dimensional region that corresponds to the current materials throughout the pre-processed image, and every independent component was unique to a particular material. The actual pre-processed images are transformed towards the source data. We choose a lot of dimensions reduced data at arbitrary, and we use the weight matrix V to determine which original group relates to the ICA conversation.

$$V_b = \frac{1}{s} \sum_{a=1}^r V_{ab} \quad (1)$$

Here, $b=1,2,\dots,k$, k =quantity of observed data, r =quantity of reduced data, and V_{ab} = how much data around the a^{th} material is contained throughout the b^{th} item.

We get a group weight series upon sorting and that series is given as

$$[V_1, \dots, V_b, \dots, V_k] \quad (2)$$

The chosen groups represent the majority of the source spectroscopic picture's qualities while retaining its material characteristics. Those groups are used to invent a fresh lower-dimensional picture.

D. Data classification using Hypothetical Naïve Bayes Algorithm (HNBA)

The HNBA seems to be the easiest probabilistic-based technique which calculates a collection of possibilities by adding the frequency with score combinations of a database. The approach is based on Bayes' hypothesis, which implies that all characteristics are separate or dependent-based upon that class variable's result. According to another description, the HNBA is indeed a classification approach that uses a probability technique using analytics to estimate future possibilities depending upon past experiences.

The possibility of every variable being calculated using the original probability. This approach could be expressed quantitatively via equation (2). Since there is no matrix multiplying or quantitative refinement, this method is generally simple for using, highly effective if used to forecast big numbers, and also has a fairly greatest accurateness throughout the prediction outcomes.

$$P(H|R) = \frac{P(R|H).P(H)}{P(R)} \quad (3)$$

Here, $P(H|R)$ =Probability with a specific location, $P(R|H)$ =under hypothesis H, the probability for variable R, $P(H)$ =preceding probability (preceding) hypothesis H, and $P(R)$ =original probability (preceding) variable R.

The HNBA approach has some features are: numerical and discrete database processing, to determine the factors (mean & deviation of factors) needed during categorization; just a minimal amount of trained information is needed, deal with the lost value via eliminating the agency when calculating the assessed possibility, quick & space-saving, and effective towards non-essential characteristics.

This approach has a few downsides also that are: never adaptable when the conditional probability was nil; if indeed the conditional probability becomes nil, then the estimated probability would also be nil, and understand there are no independent factors.

E. Prediction of student performance using the linear k-Nearest Neighbors (L-KNN)

The following are the concepts of the suggested technique: (a) Whenever these characteristics are assessed according to the circumstances of the used predictors including its present situation, the partial database, as well as weight matrix, have been acquired from KNNR; (b) linear regression approach can be used to model the objective time-series data employing the partial database as well as a weight matrix. In this section, we'll go over the L-KNN framework.

(i) Procedural structure

Let's look at the recent state of the predictors and the detected/measured datasets. This is also expected that the number of neighbors would be specified.

The necessary stages are used to model the predicted parameter:

Stage-1: Including all n measurements, we calculate the distances between both the recent as well as detected conditions of predictors.

Stage-2: Again for k-nearest distances, choose the time indexes.

Stage-3: Apply linear regression approach towards the observed datasets of the chosen time indexes.

- By employing the standard KNN selection weight, we compute the weight matrix. In opposition to the linear weight matrix, the L-KNN weight matrix isn't dependent on the recent state.
- Compute the design matrix based upon the sorted measurements we have chosen and the recent condition of the predictor factors. It's important to keep in mind that for the ordered observations, just the partial data has been used rather than the entire observation data. The weight matrix & the sorted partial data perform the same function as the multivariate kernel density estimation in the standard linear regression model in terms of distance estimation (lower distances seem to be of lower order).
- As from the weight and the design matrices, compute the variable vector through using weighted least square estimator.
- Create a model of the recent predictor.
- Stages 1 to 4 should be repeated till all the necessary data has been modeled.

(ii) Calculating the amount of closest neighbors

The quantity of closest neighbors should be correctly chosen to achieve good local linear regression estimation. The cross-validation (CV) estimation of the prediction mean squared error becomes a possible requirement (MSE).

A testing feature has been generated to incorporate the nonlinear relation with heteroscedasticity (various changes proportional to the amplitude of predictor parameters) inside the actual testing dataset to examine the effectiveness of the L-KNN framework. Additionally, the target parameter's lag-1 condition would be used as a predictor parameter.

5. Performance Analysis

In this phase, we attain the performances of the proposed system and the findings are presented by employing the MATLAB tool. Every statistical analysis of this study has been done with SPSS19.0, two-sided parametric analyses were used, relevance is set at 0.050, & $p < 0.050$ has been considered relevant.

The mean \pm standard deviation of the study findings was presented. Whenever the testing set follows a standard distribution, the double-T assessment compares among groups, while the independent sample T assessment has been used to compare groups. Two independent specimens, as well as two related specimens, would be required during the examination if indeed the standard distribution seems to be insufficient.

Figure 2 depicts that the learners' English performance has improved to some amount following training. That means prediction parameters versus certain skills. The effect of digital teaching was significantly greater than that of existing approaches. The greatest goal is caused by the distant education depending upon the L-KNN approach, which would be two times higher as before training and much more than 50percent greater than the existing instructional strategies.

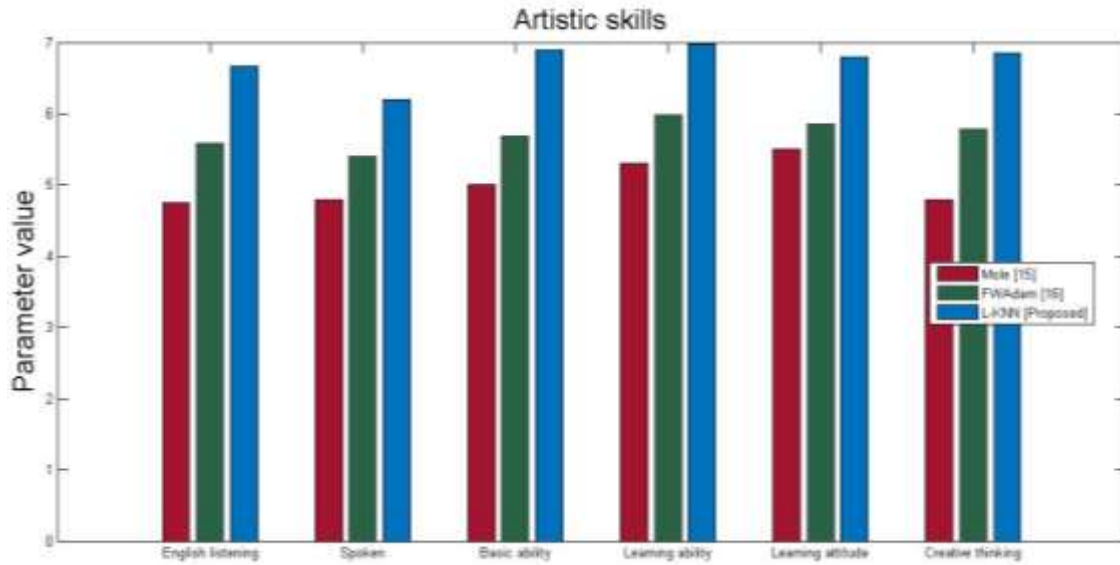


Figure 2. Comparison of artistic skills with proposed and existing approaches

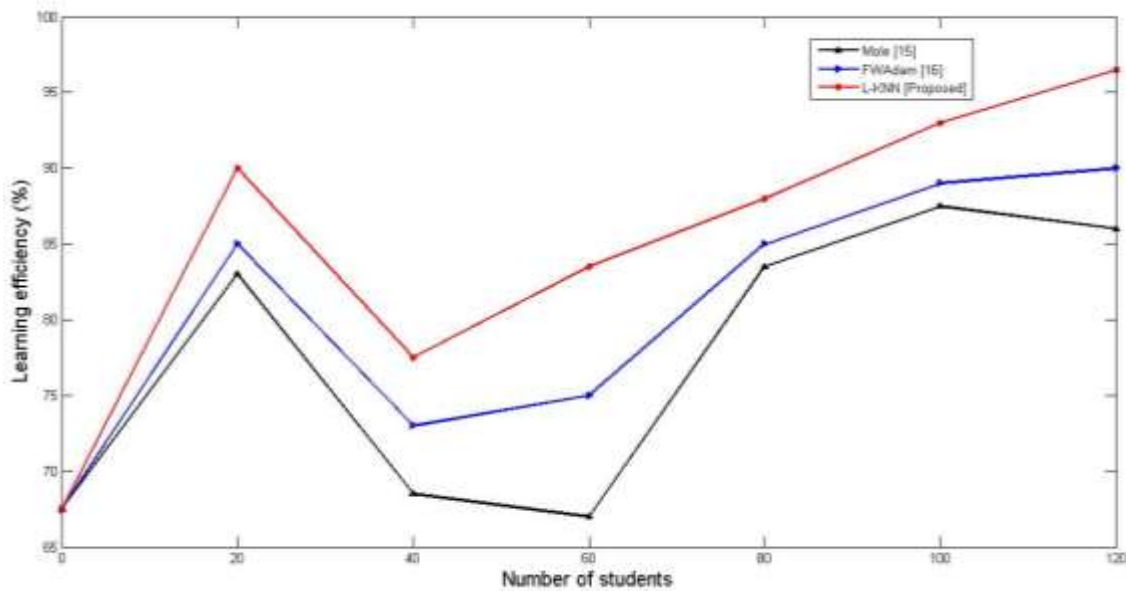


Figure 3. Comparison of learning efficiency with proposed and existing approaches

Figure 3 depicts the learning efficiency among learners. When this performance metric was matched with existing approaches, our suggested L-KNN approach seems to have higher evaluation effectiveness. It has to do with the value and efficacy of student performance in study sessions, including its significance towards the online training activity as a whole. This can be displayed in a variety of ways, including the creation and exchange of educational materials, assisting other learners, managing online discussions and answers, and having a significant impact on teams.

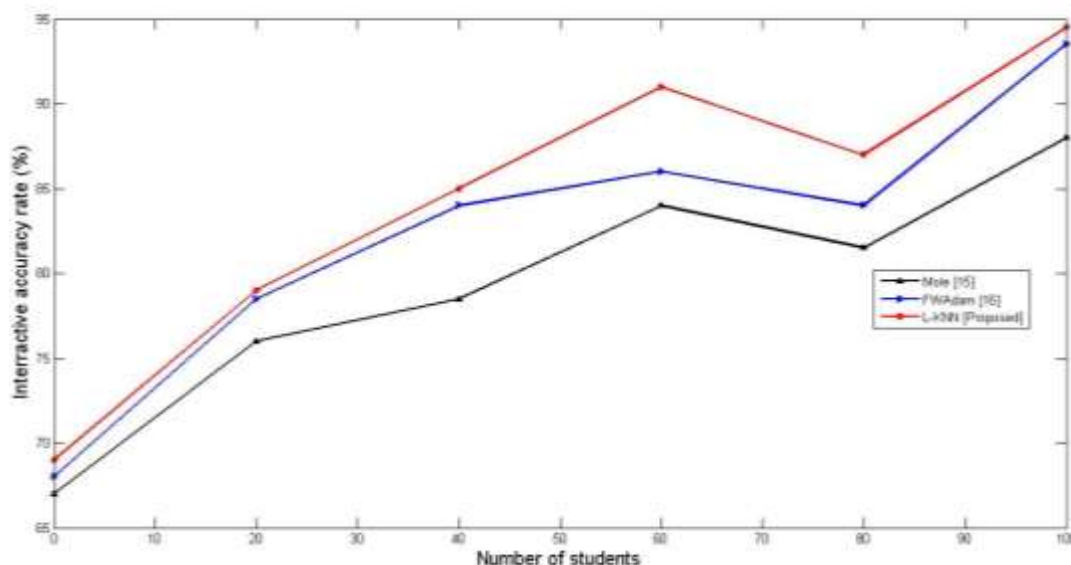


Figure 4. Comparison of interactive accuracy rate with proposed and existing approaches

As a result, if we engage successfully in a learning organization as a whole, everyone may acquire communication skills as well as receive the highest quality in English learning. Figure 4 depicts the interactive accuracy rate of our proposed approach versus students that implies the proportion of the students with interactions. When comparing our proposed approach with the existing approaches, we accomplish the greatest improvement in the interactive accuracy rate.

6. Conclusion

With the growing popularity of online teaching as well as the strengthening of online learning interactions, a huge number of structured/semi-structured and unstructured data has been gathered across diverse learning websites. There are many different forms of data. Users may examine & identify the hidden score behind it, which will help us improve our learning. This would invariably diverge from the real scenario of the learners during real-world applications. To demonstrate online classes, just a common learning system is being used. There are several restrictions. With this study, more academics will be able to focus on the diverse consequences of online learning for students. Here, we present the L-KNN approach for predicting the student's performances. Learners might further study as well as implement online learning, particularly throughout the financially disadvantaged regions, revealing its real purpose and direction.

References

- Alqahtani, S. M., Bhaskar, C. V., Vadakalur Elumalai, K., & Abumelha, M. (2018). WhatsApp: An online platform for university-level English language education. *Arab World English Journal (AWEJ)*, 9. <https://doi.org/10.31235/osf.io/j4vr9>
- Bai, H., & Zhang, Q. (2020). English smart classroom teaching system based on 5 network and internet of things. *Microprocessors and Microsystems*, 103421. <https://doi.org/10.1016/j.micpro.2020.103421>
- Cope, B., Kalantzis, M., & Searsmith, D. (2020). Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies. *Educational Philosophy and Theory*, 1-17. <https://doi.org/10.1080/00131857.2020.1824781>
- Huang, F., Lou, Y., Chen, S., & Guo, J. (2020). Students' perceptions of using mobile technologies in informal English learning during the COVID-19 epidemic: A study in Chinese rural secondary schools. *Journal of Pedagogical Research*, 4(4), 475-483. <https://doi.org/10.33902/JPR.2020063786>
- Jassim, L. L. (2020, March). Methods for learning English vocabulary employed by Iraqi EFL learners at University of Technology. In *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3798158>
- Li, X. (2017). The construction of intelligent English teaching model based on artificial intelligence. *International Journal of Emerging Technologies in Learning*, 12(12), 35-44. <https://doi.org/10.3991/ijet.v12i12.7963>

- Li, Z. (2021). Simulation of English education translation platform based on web remote embedded platform and 5G network. *Microprocessors and Microsystems*, 81, 103775. <https://doi.org/10.1016/j.micpro.2020.103775>
- Lin, L. F. (2018). Integrating the problem-based learning approach into a web-based English reading course. *Journal of Educational Computing Research*, 56(1), 105-133. <https://doi.org/10.1177/0735633117705960>
- McKinley, J. (2017). Identity construction in learning English academic writing in a Japanese university. *Journal of Asia TEFL*, 14(2), 228-243. <https://doi.org/10.18823/asiatefl.2017.14.2.2.228>
- Nagabandi, A., Finn, C., & Levine, S. (2018). *Deep online learning via meta-learning: Continual adaptation for model-based rl*. *arXiv preprint arXiv:1812.07671*.
- Palvia, S., Aeron, P., Gupta, P., Mahapatra, D., Parida, R., Rosner, R., & Sindhi, S. (2018). Online education: Worldwide status, challenges, trends, and implications. *Journal of Global Information Technology Management*, 21(4), 233-241. <https://doi.org/10.1080/1097198X.2018.1542262>
- Sun, X. (2021). 5G joint artificial intelligence technology in the innovation and reform of university English education. *Wireless Communications and Mobile Computing*, 2021, 1-10. <https://doi.org/10.1155/2021/4892064>
- Wang, P., & Qiao, S. (2020). Emerging applications of Blockchain technology on a virtual platform for English teaching and learning. *Wireless Communications and Mobile Computing*, 2020, 1-10. <https://doi.org/10.1155/2020/6623466>
- Xu, Y. (2021). Research on business English translation architecture based on artificial intelligence speech recognition and edge computing. *Wireless Communications and Mobile Computing*, 2021, 1-10. <https://doi.org/10.1155/2021/5518868>
- Zhang, M., Zhou, Y., Quan, W., Zhu, J., Zheng, R., & Wu, Q. (2020). Online learning for IoT optimization: A frank–Wolfe adam-based algorithm. *IEEE Internet of Things Journal*, 7(9), 8228-8237. <https://doi.org/10.1109/JIOT.2020.2984011>
- Zhang, X., Luo, C., He, T., Yang, X., Lu, Z., & Huang, B. (2017, October). Online learner emotional analysis based on big dataset of online learning forum. In B.M. Engineering (Ed.) and Informatics (CISP-BMEI) (pp. 1-5) 10th International Congress on Image and Signal Processing, 2017. IEEE Publications. <https://doi.org/10.1109/CISP-BMEI.2017.8302316>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).