International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I8P228

Volume 70 Issue 8, 264-271, August 2022

Original Article

Pedagogical Content Knowledge Classification using CNN with Bi-LSTM

Murthy V. S. N. Tatavarthy[1], V. Naga Lakshmi[2]

[1]*GITAM (Deemed to be University), Visakhapatnam, Andhra Pradesh, Department of Fashion Technology, NIFT Hyderabad, Telangana, India.*
[2]*Department of Computer Science, GITAM (Deemed to be University) Visakhapatnam, Andhra Pradesh, India.*

[1]tvsnmurthy@gmail.com

Received: 26 May 2022 Revised: 04 August 2022 Accepted: 17 August 2022 Published: 22 August 2022

Abstract *- Pedagogical Content Knowledge (PCK) is the establishment of academics that provides an interesting idea of teaching. PCK is an idea of the belief that teaching is considerably needed more than providing the knowledge of subject contents to students. It is also the knowledge teachers acquire over time and by the experience to explain particular contents in specific ways to students' understanding. The growth of the internet and larger digital technology has come up with various difficulties. The larger amount of data on the internet is probably unorganized and unstructured, making it difficult to utilize and manipulate the data process. Deep learning, as well as machine learning mechanisms for classifying the texts, is the importance of PCK. In the present research, the pedagogical contents are classified using Condensed Nearest Neighbor (CNN) with Bi-LSTM. In general, the CNN classifier is a simple process and constructs the subsets of example that classifies the original data correctly. The Bi-LSTM enhances LSTM, which improves the model's performance in the sequence classification process. The proposed CNN with Bi-LSTM has achieved an accuracy of 78.79compared to the existing KNN accuracy of 77.5% for the classification of pedagogical content in experiment analysis.*

Keywords *- Bi-directional long short-term memory, Condensed nearest neighbor, Digital technology, Internet, Pedagogical content knowledge.*

1. Introduction

When the teachers teach a subject, they must give its content a proper meaning and structure by considering the learning needs and prerequisites of the individual so that it will be feasible for the students to study the content. Pedagogical content knowledge (PCK) is a difficult teaching capability aspect [1]. The teaching skills have to be adapted quickly with the increasing development of science [2]. The development of teaching skills should be given more priority because, in the current era, teachers are facing tougher challenges than in the previous education era. For example, difficulties in getting used to new methods of learning activities and conditions of teaching with a higher level of standard [3]. The teachers need to follow the rules and regulations in highly complex forms and difficult subjects by learning the process standards and demands for the thinking capability of students [4]. The advantages of teaching skills in learning consist of online teaching and the delivery of courses that will not consider classroom training for students. Compared with traditional teaching methods, e-learning is cost-effective and more students can register for online classes. But, in the case of e-learning, communication between students and teachers may not be as comfortable and smooth as felt in the traditional one-to-one interactions in the classrooms [5].

So, e-learning includes some difficulties, as it is challenging for instructors to understand the course's effectiveness. Then, the student dropout rate in e-learning is higher than in the traditional learning method. Further, knowing the student's performance is difficult and predicting the risk of students in newly joined courses is also challenging. At last, the teachers will be interested in knowing the expected results of students in assessments [6]. Peer instruction is active learning of pedagogy utilized in various areas which revolve among the students by discussing and answering the concepts of questions with students and teachers. The peer instruction will be coupled with handmade devices that allow the teachers to view student answers and clickers to respond to the data collected and stored automatically [7]. The research on computational thinking capability was carried out by many researchers [8]. But the previous studies were focused highly on students and examined the thinking capability of students 10 to 16-year-old that is focused on solving the problem. However, previous researchers did not research the teachers [9]. The information provided by the existing method was ineffective for the students to decide their streams as it was not considered the influencing factors such as the subject interest, background of family, and motivation towards a career. Further, the existing systems were not provided with an acceptable outcome for higher education [10]. To resolve such issues, a pedagogical content classification method was proposed using CNN with Bi-LSTM.

The paper is arranged as it precedes; the study of existing methods based on pedagogical content classification is reviewed in Section 2. The proposed classification by CNN with Bi-LSTM for pedagogical content classification is illustrated in Section 3. In addition, the experimental outcomes and their corresponding discussion as brief in Section 4, and a conclusive summary of the proposed technique is presented in Section 5.

2. Literature Review

Several existing data analysis methods supported with pedagogical content knowledge are reviewed in this section. In this subdivision, the advantages and limitations of the existing method are detailed.

Using data mining, Mushtaq et al. [11] developed a framework for classifying educational data with pedagogical content management. This study presents a novel approach to investigating data that millions of people have already generated with a wide range of topic expertise on a shared platform. One is StackOverFlow (SO), a community Q&A website where users may pose questions and receive answers about specific issues. The developed framework classifies the stack overflow posts based on academic courses to provide learners with the best solutions. The developed data mining framework pre-processes the data by utilizing the techniques of NLP with working ML methods to classify the data. The developed framework differentiates the crowd-sourced method of education, which is suitable with the knowledge compared with an educational data management system. However, the relationship among the literature on crowd-sourced communities with the learning management system is not classified.

Madani et al. [12] developed a recommended approach for finding effective pedagogical content in e-learning using reinforcement learning. The recommendation method was based on social filtering by utilizing the notions for sentimental analysis. Youness also employed collaborative filtering to define effective ways to make the learners understand and recommend the courses better by making learners' profiles and social content. The recommendation method also consists of a reinforcement learning approach that helps identify the effective paths for learning and improves the quality of learning. But the recommendation approach utilized a similar measure which depends on inter-related items among the two users and uses overall rating information.

Zhai et al. [13] developed two assessment processes that targeted science teachers' pedagogical content knowledge, including three types of videos constructed by response questions. This research closes the communication gaps by contrasting automated and human scoring of created responses with detailed relevant data in a teacher pedagogical content knowledge (PCK) evaluation. In the developed method, the 187 science teachers watched the videos that included classroom teaching and responded with the resultant items of constructed videos. The three experts ranked the consent and responses scores of humans were utilized to establish the machine learning algorithm to predict the ratings of responses for videos. The developed machine was less delicate than the raters of humans to process the scenarios, which describes the grading to be more reconciled and stable across the scenarios among the two tasks. However, the developed methods significantly impacted the interesting construct for assessment work, and the machine was ever the severe rate.

Khoza and Biyela [14] established a pedagogical content knowledge classification method by decolonizing technology for first-year students in mathematics. This article aims to investigate and decolonize students' technological, pedagogical, and content knowledge in the study of mathematics in the first year of a bachelor's degree program. Ten individuals were purposefully chosen to participate in this research after taking a mathematics course at a South African institution. The development framed interviews, observations, and activities by the critical action of research were useful for generating data. Students' knowledge showed that the Technological Pedagogical Content Knowledge (TPACK) was effective when utilized as a framework for learning. The curriculum concepts were generated in modules to provide effectiveness for students' knowledge of pedagogy content and technology. The developed method recommends students use knowledge of pedagogy, technology, and contents in a taxonomical manner for learning.

Apuk and Nuçi [15] developed a pedagogical content classification method using LSTM with KNN. This study compares two classification methods used to categorize educational information: the KNN algorithm for the first method and the LSTM architecture for the second. The goal of the study is to understand the two methods better. Initially, text processing occurs towards exchanging unorganized information keen on a specific arrangement. Then, the features are extracted using tokenization, lemmatization, stop words removal, etc., to transform into the structured space of features.

Further, the dimension reduction is done by LDA with PCA. At last, the KNN with LSTM is utilized to classify pedagogical contents. The result of the developed method was affected because of larger similarities among various transcripts. The transcripts belonging to various classes during the course level included various similarities in keywords and sentences, so the class of transcripts was not properly distinguished.

Impact of Educational Interventional Program for Preschool Children on their Development and Awareness of sex Harassment has been reported by Amna Nagaty Aboelmagd, Ebtsam S. Mahrous, and Sabah Saleh Hassan [16]. This research adopted a pre-test and post-test quasi-experimental research design in the cities of Minia and Damanhour. This study assessed whether an educational intervention programme for preschool kids had improved the kids' knowledge and behaviour about sexual harassment. Given the secrecy and embarrassment surrounding this topic, this does not imply that there were no such events. Gender-specific reports of verbal sexual harassment varied, particularly in Cairo, where two out of every three girls experienced this type of abuse.

3. Proposed Methodology

Pedagogical Content Knowledge (PCK) is the establishment of academics that provides an interesting idea. PCK is an idea of the belief that teaching is considerably needed more than providing the knowledge of subject contents to students. This research proposed a pedagogical content knowledge classification method using CNN with Bi-LSTM. The value of PCK is realized using deep learning and machine learning techniques for text classification. The current study categorises the educational contents using CNN with Bi-LSTM. The CNN classifier creates subsets of examples that correctly classify the original data and is generally a straightforward operation. The Bi-LSTM is an LSTM augmentation that enhances the model's performance during the sequence classification phase. The proposed PCK classification block diagram is depicted in figure 1.

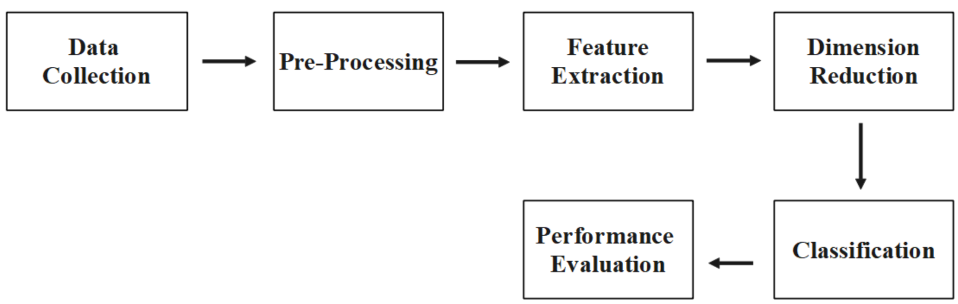

Fig. 1 The systematic diagram of proposed CNN with Bi-LSTM for pedagogical content knowledge classification.

3.1. Data Collection

The dataset used in the proposed method is employed in [17] for pedagogical content classification. The dataset includes 12,032 videos collected from the Coursera website from 200 different courses. Coursera classifies the courses among 2 levels, such as hierarchical structure from general to fine-grained level. The general level includes 8 types, the specific level includes 40 types, and the course level includes 200 categories. In addition to these three levels that made up the course, a video lesson transcript was also included.

3.2. Pre-processing

After acquiring the data, the procedure of data pre-processing takes place using natural language processing such as lexical analysis, syntax analysis, semantic analysis, pragmatics, and discourse.

Lexical analysis is the approach that analyzes the structure of sentences or words. It is the collection or set of words in a particular language to create phrases or sentences. In this method, the entire document is split into paragraphs, then into sentences, and finally, a word-based analysis is carried out. Syntactic analysis is also known as parsing or syntax analysis; it is the approach to analyzing natural language based on grammar rules. The grammar rules are applied for the groups or categories of words, and this analysis assigns the semantic structure to the text. The semantic analysis gives the process of understanding the natural language based on context and the meaning of the human interaction. The semantic analysis of the natural language process starts with reading every word in the contents to capture the text's actual meaning. Discourse integration makes one understand the meaning of any sentence based on the flow of 2nd sentence information, which is dependent on 1st sentence, which is necessary for understanding and writing the meaning of 3rd sentence flow required for reading and writing.

And in the same way, each sentence depends on the other sentence. The pragmatic analysis is the approach to retrieving information from text. Its focus is on collecting the group of text structures and representing the real meaning. The pragmatic analysis is the linguistics domain that considers the text's context.

3.3. Feature Extraction

In the accomplishment of pre-processing the pedagogical data, the procedure of pedagogical feature extraction has been accomplished using a process such as tokenization, stemming, lemmatization, stop words, capitalization, and noise removal.

Tokenization is the process of splitting corpus text into atomic words, it is a difficult approach, but it is important when punctuation marks or spaces do not delimit the words. The word replacement according to the roots or stems of tokenized words is performed by the process of stemming, and hence bag of words dimensionality is being reduced. The process of stemming involves 2 types over stemming and under stemming errors. The over stemming reduces the precision value, and under stemming decreases the recall. The whole result of stemming is based on datasets and stemming algorithms. The stop word removal process includes the connecting functions in the sentences like articles, prepositions, etc. There are no specific dictionaries for the stop words, but many searching machines utilize frequent particular words and shorter function words such as "is", "the", "at", "on", and "which". In lemmatization, the words are replaced by the root words, which include a similar meaning called lemmas. Capitalization identifies the correct capitalization of the word where the first word in the sentence will be automatically capitalized. Noise removal is removing characters, numbers, and parts of the text that affect your analysis. These characters can be some special characters, punctuation, source code removal, HTML code removal, unique characters representing a particular word, numbers, and many other identifiers.

3.4. Dimension Reduction

After extracting the features, dimensionality reduction occurs using Latent Dirichlet Allocation (LDA).

After categorizing the words and indexing the terms in the sentence, the text summarization uses LDA to extract the summary from the document. Text summarization is the method of extracting useful words from the document of texts, and it is done using LDA. The LDA is an unsupervised probabilistic algorithm that isolates the high-priority topics in a dataset described by the keywords. The documents in the datasets are considered random latent topics that provide inferred rather than observed directly. The following steps are considered to model the document by LDA.

- The whole amount of the expressions inside the manuscript is identified.
- Selected a particular topic for the document in a set of topics.
- A topic is chosen based on the document's multinomial distributions.
- Then a word is selected based on the topic's multinomial distributions.

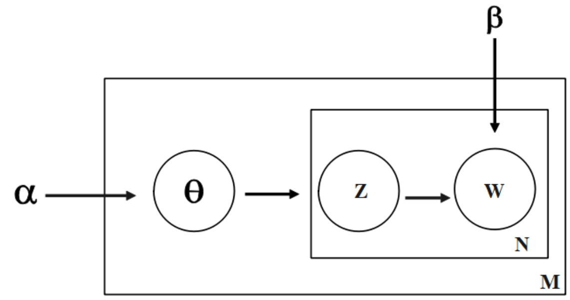

Fig. 2 Graphical representation of LDA

Figure 2 represents a graphical portrayal of the LDA model. In a document, there are vocabulary words V. Words N, documents M and latent subjects K. Each document's word w is linked to a hidden variable z representing the latent subject. A multinomial distribution is used to sample variables z, with parameters T denoting the likelihood of a latent topic. A Dirichlet distribution $K * V$ with a hyperparameter D gives the density of the T multinomial parameter. The parameter matrix represents the topic language model $\beta = \{\beta_{kw}\}$. The LDA parameters were calculated by increasing the marginal likelihood $p(\omega|\alpha, \beta)$ from a series of the collected text documents known $\omega = \{\omega_{dn}\}$, as shown in equation (1).

$$\prod_{d=1}^{M} \int P(\theta_d|\alpha)[\prod_{n=1}^{N} \sum p(\omega_{dn}|z_{dn}, \beta)p(z_{dn}|\theta_d)]d\theta_d \quad (1)$$

The latent topic named as $z = \{z_{dn}\}$ and Dirichlet parameter is denoted with θ_d regulating the marginalization. It is possible to assess LDA factors using a lower bound of (1) tends to be a sample to perform optimization using variation inference. A variational model of function $q(\theta, z|\gamma, \phi)$ is generated by evaluating the variational inference factor where the variables z are independent of optimizing the true posterior probability $p(\theta, z|\omega, \alpha, \beta)$. The various parameters $\{\gamma, \phi\}$ and the LDA parameters $\{\alpha, \beta\}$ are derived by equation to enhance the lower limit (2-5)

$$\phi_{nk} \rightarrow \beta_{kwn} \, exp\{\psi(\gamma_k)\Psi\left(\sum_{j=1}^{k} \gamma j\right)\} \quad (2)$$

$$\gamma_k = \alpha_k + \sum_{n=1}^{N} \phi_{nk} \quad (3)$$

$$\beta_{k\omega n} = \sum_{d=1}^{M} \sum_{n=1}^{N} \varphi_{dnk} \omega_{dn} \quad (4)$$

$$\alpha^{t+1} = \alpha^t - H_{LDA}(\alpha^t)^{-1} G_{LDA}(\alpha^t) \quad (5)$$

G_{LDA} and H_{LDA} are the gradient vector and Hessian matrix values of lower bound on considering α, t are represented as the iteration index of descent algorithm while ψ is known as log gamma function first-order derivative

In the decent algorithm, ψ_{it} is denoted to be the first derivative of the log function of gamma, and t is set to be the iteration index by considering H_{LDA} and G_{LDA} is the hessian matrix and gradient vector of the lower limit α, respectively.

3.5. Classification

The obtained features are forwarded to the proposed classifier model using CNN with Bi-LSTM for pedagogical content classification.

3.5.1. Condensed Nearest Neighbor (CNN)

Hart introduced the CNN algorithm in 1968 to find the subsets of labeled data points, leading to accurate and faster classification. The main limitation of the K-Nearest Neighbor (KNN) is the requirement for larger memory for storing sample data. The CNN minimizes the overall number of patterns stored in the training set of subsets. The idea behind the patterns during training is that they will be similar and add extra information and might get discarded. The CNN classifier is a simpler technique for approximating the subsets among the labeled data points. The CNN algorithm is defined with the Tgroup of labeled data points and $T(t)$ is the label identified by the t nearest neighbor classifier, which is trained T. For example, evaluating the sample subset is smaller and more accurate. The best subset Z is selected from the $2^{|s|}$ number of possible combination subsets computed. The error measure CNN is defined by using regulation theory as shown in equations (6-8):

$$Z^* = Arg\, min\, E\,(Z) \tag{6}$$

$$E(Z) = \sum_{X \in S} \quad L(x|Z) + \gamma|Z| \tag{7}$$

$$L(x|Z) = \begin{Bmatrix} 1 & if D(Z_C, x) = \left(min\, D\left(Z_j, x\right) and class(x) \neq class(Z)\right. \\ 0 & otherwise \end{Bmatrix} \tag{8}$$

Where $Z_c \in Z$ is set to be the closest stored pattern from $x \in s$ by utilizing the distance measure $D(x) \times L(x|Z)$ is the non-zero when the x and Z labels do not match.

The solutions to a problem are obtained from the theory of regularisation by combining the smoothness information and data. The initial term measures the missing data fitting because of an error obtained during classification using the nearest neighbour rule. The next term calculates the size of the subset, which is stored in such a way to define the amount of smoothness derived in the boundary separation of class. The nearest neighbor data point selected by the classifier divides the space between the input in the form of a Voronoi relation function. The class boundaries derived are piecewise linear.

3.5.2. Bi-directional Long Short-Term Memory for pedagogical content categorization

In general, Bi-LSTM is known as an RNN neural network form that represents temporal relationships. The nature of the link is direct cyclic among RNN units. It maintains the hidden neural states of the network within the internal area, in addition to aiding in the modeling of picture temporal dynamic behavior. The Bi-LSTM is a more advanced variant of the RNN, containing three gates: output, input, and forget. The Bi-LSTM is a multi-layer network with 100 nodes in each layer containing a hidden layer. Bi-LSTM establishes dependencies of the neural network's long-term temporal direction with three gates. The Bi-LSTM network efficiently addresses the gradient of vanishing problem because it reduces memory locations in temporal attributes that are often not useful in detecting final classification labels.

The function of the prior state h_{t-1} and the current word x_t is the hidden state h_t for t^{th} the word in the sub-path. As indicated in equation (9), the input is being linearly changed by the weight matrix values and squished as non-linearly as concerning the activation function (9).

$$h_t = f(W_{in}x_t + W_{rec}h_{t-1} + b_h) \tag{9}$$

Where W_{rec} are the input and recurrent connections' weight matrices, respectively. f_h is the representation of activation function of the non-linear values of layer and b_h is the representation of bias term present for the vector features in a hidden state.

Forget gate f_t output gate $o_t$and Input gate i_t memory cells i_t are the four components of an LSTM-based RNN. The three adaptive gates $i_t, f_t,\ o_t$ are all reliant on the preceding layer. The equation is used to compute the current input, and extracted features vectors are given in (10-13).

$$i_t = \sigma(W_i.\ x_t + U_i.\ h_{t-1} + b_i) \tag{10}$$

$$f_t = \sigma\left(W_f.\ x_t + U_f.\ h_{t-1} + b_f\right) \tag{11}$$

$$o_t = \sigma(W_o.\ x_t + U_o.\ h_{t-1} + b_o) \tag{12}$$

$$g_t = tanh\left(W_g.\ x_t + U_g.\ h_{t-1} + b_g\right) \tag{13}$$

Where g_t is the extracted feature gate, o_t output gate, $i_t$input gate, and f_t forget gate.

The f_t forget gate, a current recollection of cells and weighted by the input gate , provides the candidate's content g_t and gives the W weights of the particular gate as in equation (14).

$$c_t = i_t \otimes g_t + f_t \otimes C_{t-1} \tag{14}$$

Using equation (15), the hidden state's recurrent network output is calculated from LSTM Units.

$$h_t = o_t \otimes tanh(c_t) \tag{15}$$

Somewhere, the activation function uses sigmoid computation $\otimes$ and is denoted as the element-wise multiplication operator.

4. Experimental Results

The investigational outcomes of the planned pedagogical content classification model using CNN with Bi-LSTM are described. The validation of the proposed classification method carried out with the collected courser dataset against the existing approaches is described. The proposed image recovery processes are executed on the workstation through 8GB Random Access Memory by 2.2 GHz utilizing Python 3.7.3. The presentation metrics and performance investigation used for the pedagogical content categorization besides the accessible approaches explained the follows:

4.1. Performance Metrics

The estimated the presentation of the categorization process using Bi-LSTM with CNN for classifying the pedagogical content. The proposed work considered 70% of data used for training and 30% of the information used for evaluating the proposed method. The Bi-LSTM with CNN is evaluated through state-of-the-art methods and estimated using the different constraint that is utilized to check the possessions method. The performance parameters considered in the CNN with Bi-LSTM are explained as follows:

- Accuracy: Accuracy is the proportion of some predictions done correctly to the overall predictions, and it is utilized for evaluating the classification of models, as defined in equation (16).

$$Accuracy = \frac{Number of correct predictions}{overall predictions} \tag{16}$$

- Precision: The proportion of truly predicted positives to overall predictions as positive observation is termed precision and can be described with equation (17)

$$Precision = \frac{TP}{TP+FP} \tag{17}$$

- Recall: The proportion of truly forecasted optimistic explanations towards an overall number of optimistic explanations predicted are said to be recalled and can be described in equation (18)

$$Recall = \frac{TP}{TP+FN} \tag{18}$$

- F-score: F-score calculates the accuracy of the model, and it is the harmonic mean of precision and recall, which defined in equation (19)

$$F - score = \frac{TP}{TP+1/2(FP+FN)} \tag{19}$$

4.2. Quantitative Analysis

This Section explains the quantitative analysis of the proposed classification using CNN with Bi-LSTM. The values acquired for the proposed classification method utilizing CNN with Bi-LSTM for pedagogical content knowledge classification are depicted in the form of table 1. Information in Table 1 contains the evaluated results based on the pedagogical content classification in measure of parameters f-measure, class precision, accuracy, and class recall.

Table 1. The quantitative analysis of the proposed pedagogical content classification method uses CNN with Bi-LSTM.

Metrics	Proposed CNN with Bi-LSTM
Accuracy	78.79
Precision	74
Recall	76.5
F-score	75.3

Table 1 shows the quantitative investigation of Bi-LSTM with the CNN technique for classifying pedagogical content knowledge. The performances are estimated utilizing precision, accuracy, and recall with an f-score. Accuracy is the ratio of correct predictions to the overall predictions utilized for estimated method classification. The Bi-LSTM and CNN method accomplished an f-score of 75.3%, recall of 76.5%, precision of 74%, and accuracy of 78.79%. The CNN reduces the total number of patterns stored in the training set of subsets. The reasoning behind the patterns is that they will be similar during training, adding extra information that may be thrown out. The CNN classifier successfully classifies the original data and is a simpler method for approximating the subsets among the labelled data points. The Bi-LSTM is an enhancement to the LSTM that enhances a model's performance throughout the sequence classification phase. As a result, the proposed method for classifying educational content performed well in terms of accuracy, precision, recall, and F-score. The graphical cognitive content of quantitative analysis for the proposed method is shown in figure 3.

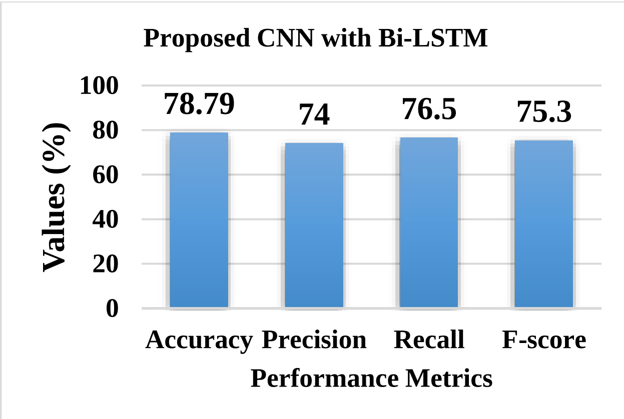

Fig. 3 The quantitative analysis graphical representation of proposed CNN with Bi-LSTM method for pedagogical content classification.

4.3. Comparative Analysis

The proposed CNN with the Bi-LSTM method for pedagogical content classification is analyzed comparatively, and the values obtained from the analysis are tabulated in table 2. The values of the existing analysis techniques, such as [12] and [15], are compared with the proposed approach. A value in Table 2 denoted evaluation investigation of technique concerning previous existing methods in terms of performance metrics such as f-score, recall, precision and accuracy.

Table 2. The comparative analysis of the proposed CNN with the Bi-LSTM method for the pedagogical content classification with the existing methods

Methods	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
Madani et al. [12]	77.0	69.0	76	67
Apuk and Nuçi [15]	77.5	73	74	72.4
Proposed CNN with Bi-LSTM	78.79	74	76.5	75.3

The proposed CNN with Bi-LSTM is compared with existing methods such as [12] and [13] for the pedagogical content classification. The existing [12] method showed accuracy of 77.0%, precision of 69.0%, recall of 76.0% and f-score of 67.0%. Similarly, in analysis, the existing [15] method represented an f-score of 75.3%. recall of 74%, the precision of 73%, and accuracy of 77.5%. The proposed CNN with the Bi-LSTM method showed an f-score of 75.3%, recall of 76.5%, precision of 74% and accuracy of 78.79%. The transcripts belonging to different classes at the course level had many similarities in the context of both sentences and keywords; hence the model could not properly distinguish which class the transcripts belonged to. In proposed CNN with Bi-LSTM method showed effective performance by combining CNN with Bi-LSTM.

Similarly, The CNN classifier is a simple process and constructs the subsets of example that classifies the original data correctly. The Bi-LSTM is the improvement of LSTM, which improves the model's performance in the sequence classification process. The graphical representation for the comparison of Bi-LSTM and CNN techniques for the pedagogical content classification is shown in figure 4.

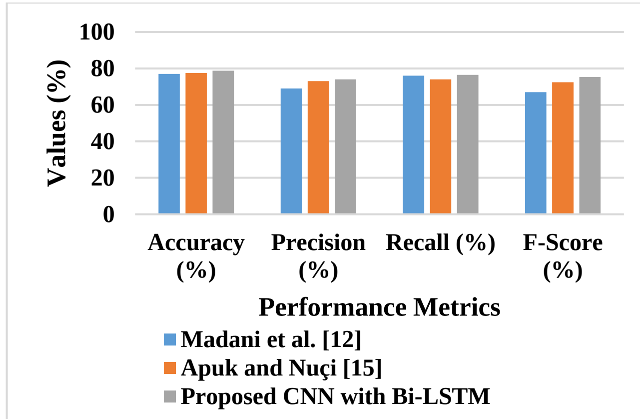

Fig. 4 The comparative graphical representation of proposed CNN with Bi-LSTM method with existing methods for pedagogical content classification.

5. Discussion

The proposed CNN with Bi-LSTM achieved an accuracy of up to 78.79% for real-time data. The reason for achieving the accuracy in this range is that CNN with Bi-LSTM is trained using a specified Coursera dataset. In contrast, the testing process is executed based on real-time data.

6. Conclusion

PCK is considered a crucial facet of teaching competence, and teaching skills need to be quickly adapted to the rapid development of education. In this research proposed, pedagogical contents of education are classified using CNN with the Bi-LSTM algorithm. The proposed method uses the Coursera dataset to achieve pedagogical content classification. The data pre-processing takes place using natural language processing such as lexical analysis, syntax analysis, semantic analysis, pragmatics and discourse. Then, the feature extraction procedure is carried out using a process such as tokenization, stemming, lemmatization, stop words, capitalization and noise removal. The obtained features are forwarded to the proposed classifier model using CNN with Bi-LSTM for pedagogical content classification. The CNN classifier is a simple process and constructs the subsets of example that classifies the original data correctly. The Bi-LSTM is the improvement of LSTM that progresses the presentation methods in the sequence classification process. The investigational outcomes illustrate that the Bi-LSTM with CNN achieved higher performance than the existing method.

References

[1] S. N. Liao, D. Zingaro, K. Thai, C. Alvarado, W. G. Griswold, and L. Porter, "A Robust Machine Learning Technique to Predict Low-Performing Students," *Acm Trans. Comput. Educ*, vol. 19, no.3, pp. 18., 2019.

[2] R. Salas-Rueda, "Construction and Evaluation of a Web Application for the Educational Process on normal Distribution Considering the Science of Data and Machine Learning," *Research In Learning Technology*, vol.27 , pp.2085, 2019.

[3] A. Wahlen, C. Kuhn, O. Zlatkin-Troitschanskaia, C. Gold, T. Zesch, and A. Horbach, "Automated Scoring of Teachers' Pedagogical Content Knowledge-A Comparison Between Human and Machine Scoring," *Front. Educ*, vol. 5 , pp.149, 2020.

[4] S. K. Fahrurozi, C. W. Budiyanto, and R. Roemintoyo, "Technological Pedagogical and Content Knowledge (Tpack) for Overcoming Teacher Problems In Vocational Education and Challenges in the 21st Century," *Journal of Mechanical Engineering and Vocational Education*, vol. 2, no.1, pp. 33–40, 2019.

[5] H. Cui, S. Tancock, and N. Dahnoun, "E-Learning tool to Enhance Technological Pedagogical Content Knowledge," *In 8th Mediterranean Conference on Embedded Computing* (Meco), pp. 1–4, 2019.

[6] J. Geng, C. S. Chai, M. S. Y. Jong, and E. T. H. Luk, Understanding the Pedagogical Potential of Interactive Spherical Video-Based Virtual Reality From the Teachers' Perspective Through the Ace Framework, Interactive Learning Environments , vol.29, no.4, pp.618–633, 2021.

[7] L. Huang, "Integrating Machine Learning to Undergraduate Engineering Curricula Through Project-Based Learning," *In Ieee Frontiers In Education Conference* (Fie), pp.1–4, 2019.

[8] K. Sethi, V. Jaiswal, and M. D. Ansari, " Machine Learning Based Support System for Students to Select Stream (Subject)," *Recent Advances In Computer Science and Communications*, vol.13, no.3, pp. 336–344, 2020.

[9] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, "Using Machine Learning to Predict Student Difficulties From Learning Session Data," *Artificial Intelligence Review*, vol. 52, no.1, pp. 381–407, 2019.

[10] S. Saarinen, S. Krishnamurthi, K. Fisler, and P. T. Wilson, "Harnessing the Wisdom of the Classes: Classsourcing and Machine Learning for Assessment Instrument Generation," *In Proc. 50th Acm Technical Symposium on Computer Science Education* (Sigcse'19), pp. 606–612, 2019.

[11] H. Mushtaq, I. Siddique, B. H. Malik, M. Ahmed, U. M. Butt, R. M. T. Ghafoor, H. Zubair, and U. Farooq, "Educational Data Classification Framework for Community Pedagogical Content Management Using Data Mining," *Int. J. Adv. Comput. Sci. Appl*, vol.10, no.1, pp.329–338, 2019.

[12] Y. Madani, H. Ezzikouri, M. Erritali, and B. Hssina, "Finding Optimal Pedagogical Content In An Adaptive E-Learning Platform Using a New Recommendation Approach and Reinforcement Learning," *J. Ambient Intell. Hum. Comput*, vol. 11, no.10, pp.3921–3936, 2020.

[13] X. Zhai, K. C. Haudek, M. A. Stuhlsatz, and C. Wilson, "Evaluation of Construct-Irrelevant Variance Yielded By Machine and Human Scoring of A Science Teacher Pck Constructed Response Assessment," *Studies In Educational Evaluation* , vol.67, pp.100916, 2020.

[14] S. B. Khoza, and A. T. Biyela, "Decolonising Technological Pedagogical Content Knowledge of First Year Mathematics Students," *Education and Information Technologies*, vol. 25, no.4, pp. 2665–2679, 2020.

[15] V. Apuk, and K. P. Nuçi, "Classification of Pedagogical Content Using Conventional Machine Learning and Deep Learning Model," *Menonet Journal: Works In Progress In Embedded Computing* (Wipiec), vol.7, no.1, pp. 1–7, 2021.

[16] Aboelmagd, Amna Nagaty, Ebtsam S. Mahrous, and Sabah Saleh Hassan, "Effect of Educational Interventional Program for Preschool Children on their Knowledge and Practice Regarding Sexual Harassment," *SSRG International Journal of Nursing and Health Science* , vol. 5, no. 2, pp. 12-19, 2019.

[17] Z. Kastrati, A. Kurti, and A. S. Imran, "Wet: Word Embedding-topic Distribution Vectors for Mooc Video Lectures Dataset,: Data Brief 28 , pp. 105090, 2020.

[18] "Boosting credibility of a Recommender System using Deep Learning Techniques - An Empirical Study", *International Journal of Engineering Trends and Technology,* vol. 69, no. 10, pp. 235-243, 2021.

[19] Ayush Jain, Prathamesh Patil, Ganesh Masud, Prof. Sunantha Krishnan, Prof. Vijaya Bharathi Jagan, "Detection of Sarcasm through Tone Analysis on video and Audio files: A Comparative Study On Ai Models Performance" *SSRG International Journal of Computer Science and Engineering* vol. 8, no. 12, pp. 1-5, 2021. *Crossref*, https://doi.org/10.14445/23488387/IJCSE-V8I12P101

[20] Dr.Ahmadi Begum, "Innovative and Effective Teaching Methods for Engineering Students," *SSRG International Journal of Humanities and Social Science*, vol. 6, no. 1, pp. 1-3, 2019. *Crossref*, https://doi.org/10.14445/23942703/IJHSS-V6I1P101