

RESEARCH ARTICLE

Multi-Agent System for Students Cognitive Assessment in E-Learning Environment

RIMSHA SHAHZAD¹, MUHAMMAD ASLAM¹, SHAHA AL-OTAIBI^{ID}², (Member, IEEE),
MUHAMMAD SAQIB JAVED¹, AMJAD REHMAN KHAN^{ID}³, (Senior Member, IEEE),
SAEED ALI BAHAJ⁴, AND TANZILA SABA³, (Senior Member, IEEE)

¹Department of Computer Science, University of Engineering and Technology, Lahore 54890, Pakistan

²Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P. O. Box 84428, Riyadh 11671, Saudi Arabia

³Artificial Intelligence and Data Analytics Laboratory (AIDA), CCIS, Prince Sultan University, Riyadh 11586, Saudi Arabia

⁴MIS Department, College of Business Administration, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia

Corresponding author: Saeed Ali Bahaj (s.bahaj@psau.edu.sa)

This research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R136), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors are also thankful to AIDA Lab, Prince Sultan University, Riyadh Saudi Arabia for support.

ABSTRACT There is a widespread increase of adoption of e-learning platforms, there is a great need to develop effective methods for assessing students' cognitive abilities to deal with that environments. Traditional assessment methods need to improve in terms of analyzing the diverse range of skills and knowledge that students acquire during online learning. The current state of research work considers different levels of Bloom's Taxonomy, which are used to evaluate students' performance based on given text analysis. For this purpose, the Software Engineering course domain is considered. An online test was conducted among graduate university level students, consisting of 12 subjective-type questions, where 2 questions chosen to be tested at each level, concluded as 12 questions per student at accumulated levels of taxonomy. There exists Approx. 300 students', who attempted the test and hence their textual responses are being used to evaluate the system. The methodology deploys the SVM classifier to predict the level of Bloom's Taxonomy for exam questions and then, multi-agent system is developed to match and identify each level assigned to an agent. Each instruction based agent is trained on Random Forest Classifier to evaluate students' cognitive skill on a scale of outcome as Good, Bad and Average. Tools, techniques, technology missing. The outcomes show that the proposed model performs much better than the existing models, provided with 98% accuracy for SVM Classifier and 92% accuracy for Random Forest Classifier to assess students' textual responses. The novelty of this research work spins around agent prediction of cognitive level for assessment by a use of multi-agent system to recognize students' strengths and weakness as per six levels of Bloom's Taxonomy.

INDEX TERMS Bloom's taxonomy, cognition, E-learning, multi-agent system, text analysis, continuous-bag-of-words (CBOW), informal learning.

I. INTRODUCTION

E-learning, which involves using a variety of technologies to assist students in finishing their courses outside of traditional/conventional classrooms, can be described as an innovative approach that is frequently employed

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan^{ID}.

in the education sector. In order to successfully develop conceptual understanding in the students, teachers must need to assess their conceptual knowledge and give timely feedback. The usage of traditional formative evaluation techniques in today's classrooms is decreasing [1]. Students' cognitive involvement can be used to examine how effective the online learning mode of education is working. Previous research has shown that student engagement significantly

affects learning as it relates to academic accomplishment [2]. Additionally, it has been stated that students who are cognitively engaged can discover new knowledge and this is necessary for pupils to learn meaning full knowledge. There are numerous techniques to assess a students' performance in an online learning platform and to evaluate their cognitive ability [23].

At several universities, e-learning has enhanced the teaching and learning processes. The type of education provided has evolved as a result, with online learning replacing traditional classroom training. In order to begin "observing" cognitive engagement, the author observed how students behaved when gathering, interpreting, summarizing, and analyzing information as well as when weighing and debating various possibilities and making decisions [3]. This was done under the premise that cognitive engagement is not elaborated in online learning. It involves looking at how questionnaires and platforms for online learning are used to assess students' cognitive engagement. The level of student cognitive involvement is a crucial indicator of how effective is the online learning. The cognitive participation of pupils is divided into three categories in this study: high, high-low, and low.

Most of the educational systems favor an online learning environment for students, and some colleges have been using virtual learning from a long time. As a result, throughout an online lecture, it is important to assess students' performance and gauge their cognitive level [4]. When examining each student's involvement through computer vision, researchers have previously either focused on their conduct (the activities they engage in) or their emotions. However, because of its increased complexity, cognition is often not addressed. Similar to this, cognitive engagement is related to students' mental efforts when they may show a high level of knowledge and improved learning [5]. Furthermore, it assessed the cognitive domain in online forums using computerized text analysis.

To measure the amount of students' cognitive involvement in virtual learning, this study, however, used levels of Bloom's Taxonomy. Bloom held that one must first primary the first level in order to advance to the second level and beyond. The pyramid's lowest level represents the simplest form of learning, while the top level represents the most challenging and abstract form of cognitive aptitude as shown in Figure 1.

Below [29] is a quick explanation of each level:

1. **Knowledge:** At this level, students must demonstrate retention of previously studied content by recalling facts, conditions, key concepts, and solutions. This level is referred to as "recalling of learning material" and is easier. Example keywords required for identification of first level.
2. **Comprehension:** Students at this level must classify, contrast, translate, grasp, describe, and condition essential ideas to show that they understand facts and concepts.
3. **Application:** One needs to acquire new knowledge in order to fulfill the requirements for this level.

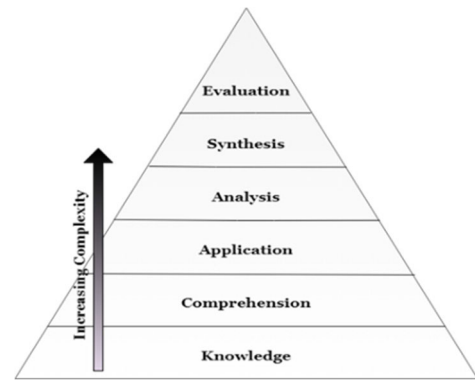


FIGURE 1. Fixed hierarchy of bloom's taxonomy levels.

By using recently gained knowledge, facts, methods, and policies, one must deal with issues in novel settings. Applied side example, Applications of Usability heuristics in mobile applications.

4. **Analysis:** Students at this level are likely to segment the material by establishing objectives, justifications, or conclusions and to collect data to back up generalizations. comparative analysis among technologies.
5. **Synthesis:** At this stage, each student individually gathers information in a distinctive way by putting the fundamentals together to create a new model or putting forth alternatives. Hypothesis and logical reasoning to state the answer.
6. **Evaluation:** In order to express and defend viewpoints, students must develop evaluations of knowledge, the talent of ideas, or the excellence of work based on the set criteria.

For a more accurate assessment during an online exam for theoretical courses, there should be an appropriate technique to map cognitive level with Bloom's Taxonomy through text analysis. In order to obtain precise results, we introduced a cognitive evaluation technique for evaluating students' performance [6]. The cognitive evaluation models that have already been created were either based on the physical classroom setting for student presence pertinent to eLearning platform. However, the goal of the system is to assess students' cognitive abilities in order to guarantee their participation in achieving learning objectives [7]. The derived results are then blended as direct proportional to measure the students' eventual tendency to study in an eLearning environment [27]. This study's primary goal is to determine how well students do in an online learning environment using a content analysis paradigm for cognitive evaluation. To more accurately measure student achievement, the system must build a new method involving a multi-agent system in which each agent must align to a particular level of Bloom's Taxonomy [31]. To identify students' strengths and limitations at each learning level, the established system must give feedback to mentors for the evaluation of students' mental states [22].

The remaining sections of the study paper are organized as follows. Section III includes a comprehensive literature assessment of previous studies. Section IV presents the suggested technique and explains the architectural concept with respect to relevant aspects in the domain of cognitive assessment. Experiments and experimental setup used to test the proposed framework are presented in Section V. Results and discussion are presented in Sections VI and VI, respectively. Section 8 is the last section of the paper giving conclusion.

II. LITERATURE REVIEW

Recent research has focused on measuring students' cognitive abilities in e-learning environments using baselined methodologies etc. Since a few decades ago, a lot of researchers have been concentrating on the cognitive domain. Content analysis is mostly used to identify cognitive level and link it to students' performance on online tests. Through a social knowledge creation coding method, the scripts from discussion forum files and server log files are collected and assessed. The author also emphasized that when texting their views in a discussion forum, students frequently did not engage their thinking abilities [1]. They provided a model with four basic steps: corpus development, feature extraction and selection, classification strategy, and testing. The cognitive levels of the communication are determined using that model [3]. Additionally, a Natural Language Processing (NLP) technique was used to automate text analysis-based assessments of students' conceptual knowledge. From this pattern, features are taken. The next step is to employ three models as classifiers to determine mental state, including SVM, Linear Discriminant Analysis, and K-Nearest Neighbor. SVM's superior accuracy for this type of detection has also been emphasized [13].

Another study used a strategy of employing questions that were automatically created to test the real-time cognitive characteristics of online learners. To do this, text from video lectures is gathered, questions are created using the model, and responses are then evaluated [6]. The Bloom's Taxonomy is a useful tool for categorizing cognitive abilities. To investigate that, several researchers have worked in this area. By the inclusion and deployment of regime of Bloom Taxonomy, assess students' cognitive abilities in online classes. Students completed an objective question paper from the Software Engineering course for this purpose, and their answers were assessed using Bloom's Taxonomy's six levels [24].

Additionally, it has been noted that this kind of testing is beneficial for identifying kids' learning strengths and weaknesses. However, it was discovered that a number of the action verbs in Bloom's Taxonomy overlap at various levels of the hierarchy, creating uncertainty regarding the precise amount of cognition required. Two methods are utilized in this study to identify the cognitive complexity of 2000+ questions [20]. When applying the BERT framework for text-classification, we were able to attain 89% accuracy compared to the LDA model's 81% accuracy. It found a link between students' academic success in eLearning and their

motivation levels. Additionally, 111 social science students participated in a motivating survey, and data was analyzed using SPSS software.

The findings showed that there was a positive but tenuous relationship between academic achievement and encouragement level [8]. Determined cognitive load (CL) based on assessing students' performance, attentiveness, behavioral, and subjective characteristics with the goal of measuring cognitive load. Distractions [9]. Students' programming code collects data, and method is then used to automatically match the students' competency level with the pertinent cognitive level from the written code. The findings of the model evaluation and the manual assessment are then compared [10]. Researchers have created a novel method for categorizing students' cognitive domain evaluations of free body diagram subjects in physics. To do this, the test is created so that the only possible answers are diagrams, and the students' responses are then analyzed in accordance with their RBT-based cognitive levels.

Some researchers suggested a prediction model to investigate students' intellectual participation in distant learning. Students' engagement is tracked by LMS data, and their written remarks during problem solving exercises are also gathered [11]. Finally, written messages are assessed using a cognitive involvement coding structure and student involvement is analyzed using sign in activity, how frequently they are accessing course materials and debates.

As can be seen in Table 1, the majority of scholars have previously conducted work related to the evaluation of students' cognitive abilities in a distant learning setting. The following table summarizes the most relevant articles that are carefully chosen as shown in Table 1.

In research, conceptual knowledge is often extracted from textual answers using NLP and neural network algorithms. The primary goal of the proposed study is to investigate the link between cognitive engagement and academic success in e-learning and to develop a multi-agent system for a more accurate evaluation of students' cognitive ability [33].

III. PROPOSED METHODOLOGY

To improve the effectiveness and efficiency of evaluating students' cognitive capacities in an online learning environment, a Multi-Agent System has been developed. Agents based System is a computational system that consists of multiple autonomous agents that interact and collaborate to achieve a common goal. In the context of cognitive assessment, Multi-Agent System can provide personalized assessment and feedback, adapt to individual learning styles, and support the overall assessment process. Additionally, each agent can measure students' cognitive performance in an online learning environment via their textual answers, after transforming the text into numerical features.

A. SYSTEM ARCHITECTURE

The primary functionalities of the suggested system are indicated by the component diagram of the proposed technique

TABLE 1. Comparison with latest state of the art.

Year - Ref. No	Methodology	Outcomes	Participants	Research Gap
2023 - [15]	RoBERTa-CNN model for Learner's cognitive evaluation	Overall accuracy is 85.26%	Text data from MOOC	No testing on larger dataset
2023 - [16]	Regression model to classify subjective questions	Prediction of Bloom's Taxonomy Level	Item Bank – Computer Networks	No testing on additional course questions
2022 – [17]	Classifying Questions using Text Mining method	Questions analysis using Bloom's Taxonomy	Exam Questions	Need to map cognitive level for answers
2021 – [4]	NLP technique to assess students' performance via text analysis	Accuracy is 93%	Course – Signal Analysis	Need to map cognitive levels
2021 – [7]	Students' evaluation using Bloom's Taxonomy	Useful to assess performance	Course – Software Engineering	Need to work on subjective type questions
2020 – [13]	Classification of cognitive skill based on RBT and KNN	Overall accuracy is 84%	Computer Science students'	Need to compare other NLP models
2020 – [14]	RBT, SVM and K-Mean Clustering to assess students'	Accuracy is 97%	Bachelor students'	No textual response analysis
2019 – [11]	Utilization of RBT Levels on Physics Diagram	HOTS and LOTS assessment	Physics course	No testing on larger dataset

in Figure 2. Using Bloom's Taxonomy, the system creates a performance forecast model to categories the cognitive levels. All six levels' worth of questions based on Bloom's Taxonomy are fed into the model as input. Each question is parsed by the model by tokenizing each word, and stop words are then eliminated from the question using lemmatization methods. Additionally, the model employs stemming techniques to reduce the most important terms in the query to their most basic form. Finally, the questions are transformed into a dataset with the significant terms acting as independent variables and the question's Bloom's Taxonomy level acting as a class or dependent variable. The support vector machine classifier algorithm, which is extensively used and frequently yields the best prediction accuracy compared to any other classifiers available in the research, is utilized in the model as shown in Figure no 2 below.

The methodology consists of various stages, including design of sample question paper, data collection, data pre-processing, and Bloom's Taxonomy level prediction model, design of agents for student's textual response analysis, testing and evaluation of proposed system. In order to train SVM Classifier, sample software engineering course questions are gathered from previous research work and via web scraping from various online websites; such as GeeksforGeeks, Guru99, TutorialsPoint and IndiaBix etc. After that, questions preprocessing is done using Natural Language Toolkit library

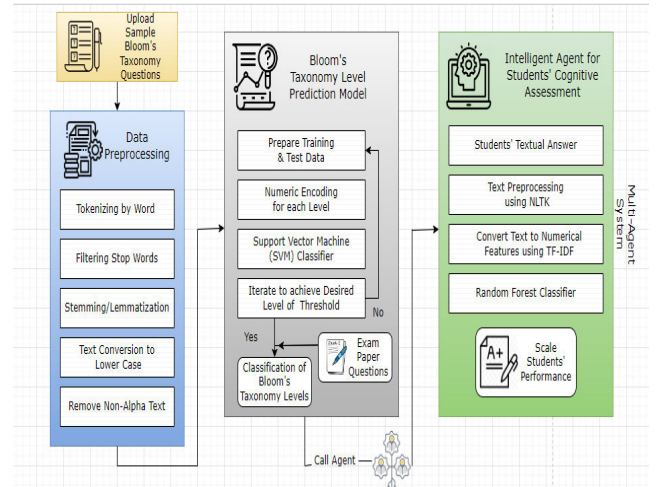


FIGURE 2. High level view of proposed methodology.

in Python and numeric encoding is assigned for Bloom's Taxonomy Level. Each question is converted into feature vectors using TF-IDF and Support Vector Machine (SVM) is trained on this dataset to classify questions into cognitive levels. Once SVM classifier is ready, multi-agent system is designed using python to train each agent for textual response analysis. Dataset for textual answers is collected using Beautiful Soup Library and then, text features are extracted using TF-IDF and Word2Vec to train NLP model such as Random Forest Classifier on each agent. Once the system is ready for use, dataset of collected students' responses is provided to scale cognitive level into Good, Bad and Average for the evaluation of students' performance at each level.

B. SYSTEM MODULES

Following is the detailed explanation of each module.

1) DATA PREPROCESSING

This step involves the use of NLP approaches for pre-processing of raw textual data. Exam questions and students' textual answers are pre-processed to convert them into the form of dataset. Stop words, slang words, misspelled words etc. are common in textual answer and these extraneous characteristics can have a negative impact on the performance of ML based algorithms. Therefore, it is necessary to do text pre-processing first using NLTK as shown in Figure 3.

2) FEATURES EXTRACTION

To transform textual information into numerical vectors usable by machine learning models, text feature extraction is a popular approach in NLP. TF-IDF and Word2Vec are two well-known methods for extracting text features.

The Term Frequency- Inverse Document Frequency (TF-IDF) is a measure of a word's significance in a collection of words as a numerical statistic. In NLP, TF-IDF is a popular method for determining relationships between words.

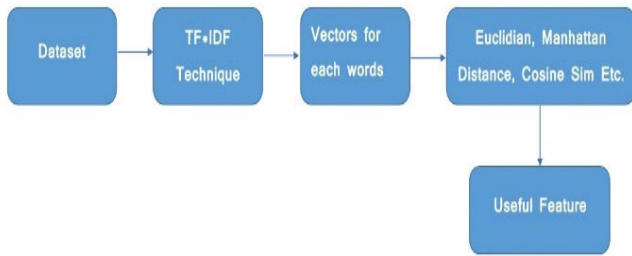


FIGURE 3. Textual answer pre-processing.

Words that are frequent in a text but uncommon in the textual answer are assumed to have more informational value in TF-IDF. This method assigns each word a score based on how often it appears in the text. The mathematical representation of its formula is given below in Eq. 1:

$$W(d, t) = TF(d, t) * \log(N/df(t)) \tag{1}$$

As the name implies, TF-IDF vector scores a word by multiplying the word’s Term Frequency (TF) with the Inverse Document Frequency (IDF). TF stands for Term Frequency, where the term or a word is the number of occurrence’s in a documented dataset, as compared to the total number of words in the document.

Thus, features from text data can be extracted using TF-IDF as shown in Figure 4 and then, can be utilized in ML models for tasks like classification and clustering. A machine learning algorithm can feed a feature matrix that contains the TF-IDF score for each term in a text.

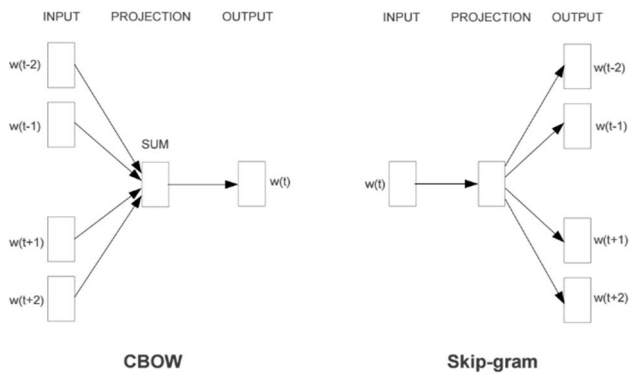


FIGURE 4. TF-IDF features extraction.

Words are often represented as vector using “Word to Vector” method. We normally use that technique to figure out words semantic. Word2Vec uses a neural network based architecture to learn from a vast body of textual answer. The neural network is fed with text to make probabilistic predictions about individual words in that text based on their surrounding context. This technique will represent each term as vector and there will be a close connection between the vectors, representing words that have equivalent meanings. For example, vectors for “software” and “engineering” will

be near together but, the vectors for “software” and “water” will be far away.

```

Student Answer: Software Engineering is a field of computer science to study softwares.
Step1: Tokenization
['Software', 'Engineering', 'is', 'a', 'field', 'of', 'computer', 'science', 'to', 'study', 'softwares', '.']
Step2: Lower Case Conversion
['software', 'engineering', 'is', 'a', 'field', 'of', 'computer', 'science', 'to', 'study', 'softwares', '.']
Step3: Remove Stop Words
['software', 'engineering', 'is', 'a', 'field', 'of', 'computer', 'science', 'to', 'study', 'softwares']
Step4: Stemming
['softwan', 'engin', 'is', 'a', 'field', 'of', 'comput', 'scienc', 'to', 'studi', 'softwan']
Step5: Remove Special Characters
['softwan', 'engin', 'is', 'a', 'field', 'of', 'comput', 'scienc', 'to', 'studi', 'softwan']
  
```

FIGURE 5. Word2Vec techniques.

Both the Continuous-Bag-of-Words (CBOW) and the Skip-Gram models can be used to train the Word2Vec model as shown in Figure 5. The current word can be predicted by a CBOW architecture, while neighboring words from the provided current word can be predicted using the Skip-Gram model. In our developed system, we used the CBOW model to implement the Word2Vec technique that can be combined with features extracted from previous method - TF-IDF to get more accurate results.

3) BLOOM’S TAXONOMY LEVEL PREDICTION MODEL

After text pre-processing and features extraction, next module is Bloom’s Taxonomy Level prediction model. This step is mainly used to classify questions on the basis of cognitive levels so that, relevant agent can be called for textual response analysis in next module.

For the prediction of Question level as per Bloom’s Taxonomy, SVM Classifier is used which takes numeric features of questions as an input and produce output in the form of numeric encoding for cognitive levels. Each level is assigned a numeric encoding in the sequence of 1 to 6.

To classify data, SVM seeks for a hyperplane in a feature space with high dimensions that best divides the data into distinct categories. A hyperplane breaks down a decision boundary between two sets of data. In order to maximize the separation between the various Bloom Taxonomy Level classes during training, the SVM algorithm finds the optimal combination of support vectors (data points nearest to the decision border). In this case, the radial basis function has been as a kernel.

4) STUDENTS’ COGNITIVE ASSESSMENT MODEL

Once question level is identified by previous module, next step is to call relevant agent for text analysis and to assess students’ performance on that question. If a question is classified as Knowledge level then, knowledge level agent would be called for the analysis of textual response and same goes for all other agents. Total six agents have been designed and trained on Random Forest Classifier. Each agent will get textual response features, extracted from TF-IDF and Word2Vec as an input and will produce outcome on

a scale of good, bad and average to categorize students' performance. Description is required for student's ultimate classification.

Following values are being used to set model's parameters: $n_estimators = 600$, which means 600 trees were evaluated; $max_depth = 4$, which means the maximum depth for each tree was evaluated; $max_features = 3$, which means no more than three features were considered for inclusion in any given tree; $bootstrap = True$ (the default; to emphasize how bootstrapping is relevant to random forest models); and $random_state = 18$, default value set. Once RF models is trained on each specific agent using textual answers of Software Engineering course questions, it represents student performance at different taxonomy levels as shown in Figure 6.

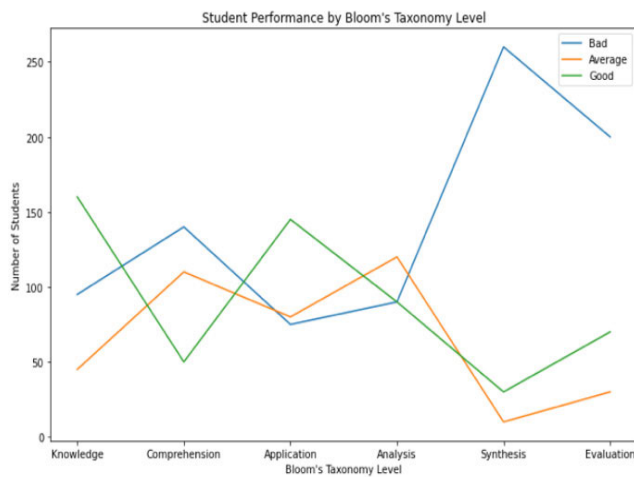


FIGURE 6. Random forest classifier prediction.

The students' performance is tested, based on students' provided responses, in order to evaluate engagement in online learning platform.

IV. EXPERIMENTATION

This section describes the setup of our experiments which includes the main libraries used along with dataset collection. Figure 7 shows the accuracy comparisons of different algorithms from cosine similarity gives 70%, Naïve Bayes 92% and 98% for SVM.

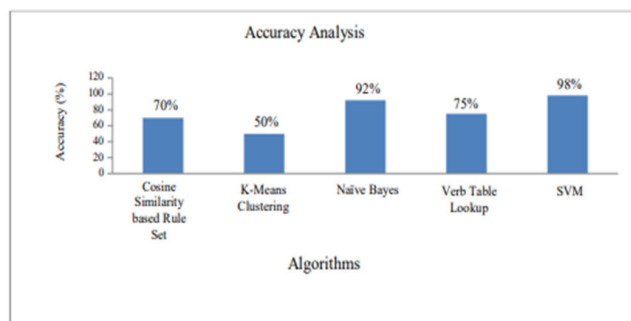


FIGURE 7. Comparison of SVM classifier.

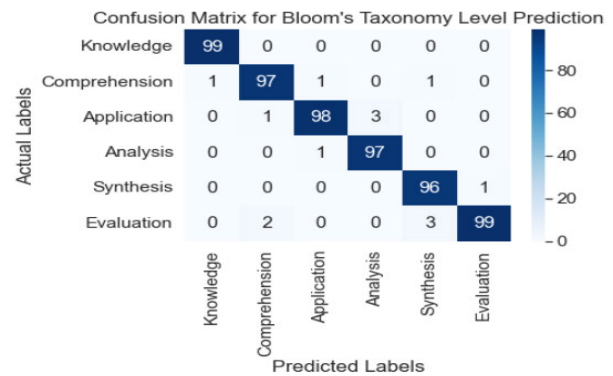


FIGURE 8. Confusion matrix of SVM classifier.

A. EXPERIMENTAL SETUP

Google Collab and the Python programming language is being utilized to conduct the experiment. When it comes to deep learning and machine learning, Google Collab is a popular online IDE. Moreover, the following libraries are being used throughout the experiment's execution. There is an 80:20 split between the dataset's training and testing sets.

Different libraries related to NLP are used for text analysis such as Python 3, Pandas, Requests, BeautifulSoup, Natural Language Toolkit (NLTK), Numpy, Scikit-learn, Gensim, Matplotlib and Seaborn.

B. DATASET COLLECTION

The use of Bloom's taxonomy in teaching has been widely adopted. In this study, an online software engineering course is taken into account for student evaluation. To meet the learning objectives of the software engineering course and to guarantee that the assessment is completed effectively, assessors must categories questions based on various cognitive levels [7]. For this purpose, a Google Form is used to create an online test that asks 12 subjective questions, each one corresponding to one of the six levels of Bloom's Taxonomy, in order to measure the mental condition of students'. An online test is conducted among students enrolled in the fifth semester of the Bachelor of Computer Science - Software Engineering Course. It will be transformed into a graded activity to ensure maximum student input and the exam will be administered at the conclusion of the online class in order to motivate students for this assessment and to accomplish our purpose. The online test has a total of 12 questions, as was previously discussed, and the allotted time is 30 minutes. The collected dataset includes 300 students' textual responses, and it serves as an input to an intelligent system.

The criterion established to determine the scale of students on the metric of Good, Bad and Average depends upon their scores at each level, we have student with $>60\%$ in Analysis and $>70\%$ in Application is considered as at Good in Cog levels, similarly the student with Analysis level score $<60\%$ and $>50\%$ score in Application is considered as below Average and Average.

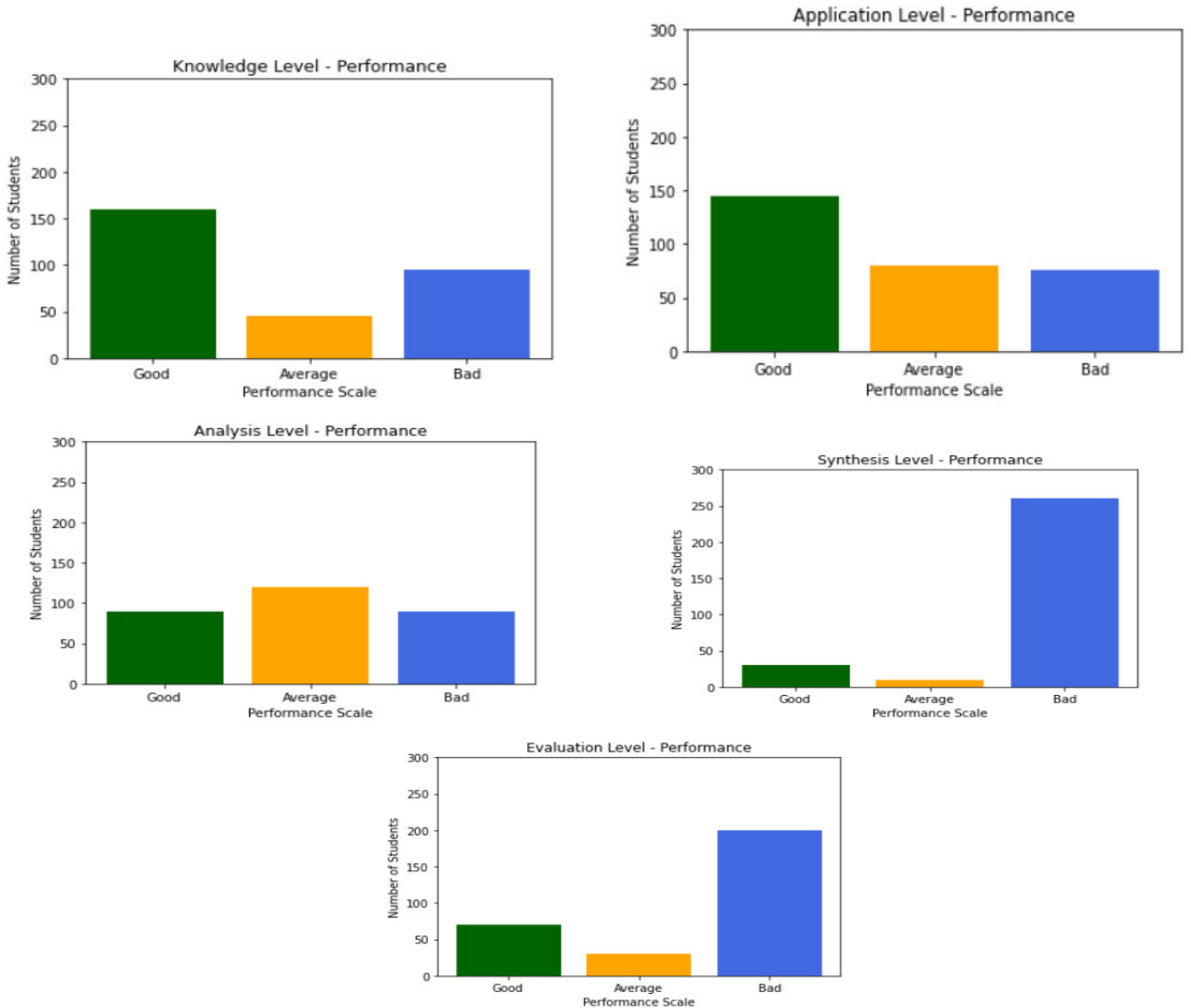


FIGURE 9. (A). Students’ performance results. (B). Students’ performance results. (C). Students’ performance results. (D). Students’ performance results. (E). Students’ performance results.

V. RESULTS AND DISCUSSION

The findings of this study will be discussed below. Since there are two main phases of methodology in our developed system, the results are also divided in two sections as follows:

A. SVM CLASSIFIER RESULT

The questions in online tests are designed according to Bloom Taxonomy level and SVM classifier is used for the classification of Bloom Taxonomy levels. We have chosen SVM for classification because according to recent literature, it is the most widely used classifier.

Additionally, SVM classifier achieves the best result with respect to other machine learning based algorithms for predicting the accuracy of Bloom Taxonomy level classification, as observed in Figure 8.

The confusion matrix provides an additional metric for evaluating the performance of the proposed SVM classifier.

The confusion matrix compares the proportion of cases that are properly categorized by the model against the proportion that are incorrectly predicted. The diagonal values represent the number of correctly classified labels which are true in reality. The SVM classifier’s confusion matrix has been shown in Figure 8. The values in the confusion matrix is scaled up values. The actual instances are small therefore all the instances and the outcomes of model are scaled up out of 100.

The performance of the SVM classifier is also measured in terms of accuracy and kappa statistics as observed in Table 2.

B. RANDOM FOREST CLASSIFIER RESULT

After question level prediction, numeric features of textual response from TF-IDF and Word2Vec is provided as an input to respective agent and then, machine learning model is being

TABLE 2. SVM classifier results.

Measure	Value
Accuracy	97%
Kappa	0.973

used to provide outcome as a performance scale. Random Forest is one of best classifier for text classification to evaluate students' performance. Each agent returns assessment results in the form of good, bad and average per 300 students' and then, results are combined to find out overall accuracy of the system as shown below in Table 3.

TABLE 3. Combined experimental results of RF model.

Model	Indicator	Bloom's Cognitive Level					
		Level1	Level2	Level3	Level4	Level5	Level6
RF	Precision	92.81	86.23	87.32	79.43	84.01	90.86
	Recall	87.00	89.42	84.21	89.01	89.98	91.12
	F1	89.34	88.03	85.09	90.23	92.86	95.23
	Accuracy	91.83					

C. STUDENTS PERFORMANCE EVALUATION

The analysis of results shown that how well student performed in each level of Bloom's Taxonomy, as observed in Figure 9. By considering all levels into a single performance scale, there are 144 students who performed badly in all the Bloom Taxonomy levels, 66 students have shown average performance and 90 students performed well in online test. The overall classified performance of blooms taxonomy level in shown below in different graph level, shows Knowledge level of student performance is represented vat Good level as per Figure 9 (A).

The Performance as per Application level is shown as Good in the cognitive level testing of students as shown in Figure 9 (B)

The Performance as per Analysis level is shown as Average in the cognitive level testing of students as shown in Figure 9 (C)

The students' performance at the difficult level of synthesis shows students unsatisfactory performance and it serves as an input for students to improve their performance and learn for the improvement to compete at the synthesis level of taxonomy as well as compared to the other cognitive levels performance, as shown in Figure 9 (D).

The students' performance at evaluation level shows majorly the unsatisfactory result due to the complexity of the cognitive level as shown in Figure 9 (E).

The Table 3 shows the aggregate overall accuracy of 91.83 as per students' performance at different levels of blooms taxonomy sums over precision, recall, F1 and Accuracy parameters.

To find out the students who performed badly in Bloom Taxonomy level, the experiment is repeated again in an episodic way. Figure 9 shows the bar chart representation of students' performance in software engineering course.

Students' strengths and weaknesses can be more easily seen when their progress on every level of Bloom's Taxonomy is visualized using a bar chart. According to results, most students did well on the knowledge and application questions, but only few of them did well on the synthesis and evaluation questions. This shows that teachers should place a greater emphasis on helping students acquire higher-order thinking abilities.

VI. CONCLUSION

This research work provides a solution to the problem of evaluating student's performance in an online environment. In this study, different levels of Bloom's Taxonomy are used to measure the students' cognitive level and multi-agent's system is developed to evaluate students' performance. Each agent is mapped to one level of Bloom's Taxonomy. This study has the novelty that it covers questions which are subjective in nature. Previous studies are limited to objective type questions only. Different NLP techniques are used to process the data, gathered from students' enrolled in Software Engineering Course. The processed dataset is then used to predict the Bloom's Taxonomy level of question using SVM classifier with 98% accuracy. After identification of the level, the corresponding level agent is called to analyze students' textual answer. For this purpose, the Random Forest algorithm classifies students' performance on a scale of bad, average and good, whose accuracy is 92%. The results showed that the proposed methodology is able to achieve results, which are better than the existing studies. The system can be enhanced to evaluate students' cognitive skill in other courses or in SE- Lab assignments. Also, the system can be modified to provide real-time feedback to students. Agents can be trained to compare the student responses with the actual answers and can point out the differences in answers. In future, the system can also be enhanced in such a way that it can predict students' mental state by assessing the grades in all the subjects, not only in one subject.

DECLARATION

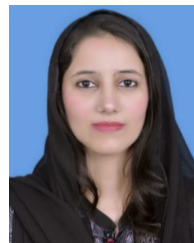
The authors declare no conflict for this research.

ACKNOWLEDGMENT

This research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R136), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors are also thankful to AIDA Lab. Prince Sultan University, Riyadh Saudi Arabia for support.

REFERENCES

- [1] S. N. Kew and Z. Tasir, "Analysing students' cognitive engagement in e-learning discussion forums through content analysis," *Knowl. Manag. E-Learn.*, vol. 13, no. 1, pp. 39–57, 2021.
- [2] P. Vanneste, J. Oramas, T. Verelst, T. Tuytelaars, A. Raes, F. Depaepe, and W. Van den Noortgate, "Computer vision and human behaviour, emotion and cognition detection: A use case on student engagement," *Mathematics*, vol. 9, no. 3, p. 287, Feb. 2021.
- [3] A. Tenriawaru and A. Rahman, "An automated cognitive analysis model for online asynchronous discussion," *J. Phys., Conf. Ser.*, vol. 1899, no. 1, May 2021, Art. no. 012118.
- [4] R. Somers, S. Cunningham-Nelson, and W. Boles, "Applying natural language processing to automatically assess student conceptual understanding from textual responses," *Australas. J. Educ. Technol.*, vol. 37, no. 5, pp. 98–115, Dec. 2021.
- [5] T. A. Suhail, K. P. Indiradevi, E. M. Suhara, P. A. Suresh, and A. Anitha, "Electroencephalography based detection of cognitive state during learning tasks: An extensive approach," *Cognition, Brain, Behav. Interdiscipl. J.*, vol. 25, no. 2, pp. 157–178, Jun. 2021.
- [6] R. Gala, R. Vijayaraghavan, V. Nikam, and A. Kiwelekar, "Real-time cognitive evaluation of online learners through automatically generated questions," NASA/ ADS Cornell Univ., Tech. Rep., 2021.
- [7] G. N. R. Prasad, "Evaluating student performance based on Bloom's taxonomy levels," *J. Phys., Conf. Ser.*, vol. 1797, no. 1, Feb. 2021, Art. no. 012063.
- [8] K. S. Na, S. Petsangsri, and Z. Tasir, "The relationship between academic performance and motivation level in e-learning among Thailand University students," *Int. J. Inf. Educ. Technol.*, vol. 10, no. 3, pp. 181–185, 2020.
- [9] N. Herbig, T. Düwel, M. Helali, L. Eckhart, P. Schuck, S. Choudhury, and A. Krüger, "Investigating multi-modal measures for cognitive load detection in E-learning," in *Proc. 28th ACM Conf. User Modeling, Adaptation Personalization*, Genoa, Italy, Jul. 2020, pp. 88–97.
- [10] Z. Ullah, A. Lajis, M. Jamjoom, A. H. Altalhi, J. Shah, and F. Saleem, "A rule-based method for cognitive competency assessment in computer programming using Bloom's taxonomy," *IEEE Access*, vol. 7, pp. 64663–64675, 2019.
- [11] C. Poluakan, A. F. Tilaar, P. Tuerah, and A. Mondolang, "Implementation of the revised Bloom taxonomy in assessment of physics learning," Phys. Departement, Manado State Univ., Manado, Indonesia, Tech. Rep., 2019.
- [12] N. Süzen, A. N. Gorban, J. Levesley, and E. M. Mirkes, "Automatic short answer grading and feedback using text mining methods," *Proc. Comput. Sci.*, vol. 169, pp. 726–743, Jan. 2020.
- [13] C. Troussas, A. Krouska, F. Giannakas, and I. Voyiatzis, "Automated reasoning of learners' cognitive states using classification analysis," *PCI*, vol. 2020, pp. 103–106, Nov. 2020.
- [14] J. C. A. Sami and U. Arumugam, "A descriptive analysis of students learning skills using Bloom's revised taxonomy," *J. Comput. Sci.*, vol. 16, no. 2, pp. 183–193, Feb. 2020.
- [15] Y. Cai, Y. Zhao, H. Luo, Z. Huang, Z. Huang, and F. Zhang, "Evaluation model of Learner's cognitive level based on RoBERTa fused with CNN," in *Proc. 3rd Int. Conf. Neural Netw., Inf. Commun. Eng. (NNICE)*, Guangzhou, China, Feb. 2023, pp. 66–69.
- [16] M. O. Gani, R. K. Ayyasamy, A. Sangodiah, and Y. T. Fui, "Bloom's taxonomy-based exam question classification: The outcome of CNN and optimal pre-trained word embedding technique," *Educ. Inf. Technol.*, vol. 28, no. 12, pp. 15893–15914, Dec. 2023, doi: 10.1007/s10639-023-11842-1.
- [17] R. Lourdusamy, P. Magendiran, and C. M. Foncoca, "Analysis of cognitive levels of questions with Bloom's taxonomy: A case study," *Int. J. Softw. Innov.*, vol. 10, no. 1, pp. 1–22, Jun. 2022, doi: 10.4018/ijsi.297922.
- [18] N. Ghatasheh, "Knowledge level assessment in e-learning systems using machine learning and user activity analysis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 4, pp. 107–113, 2015.
- [19] N. A. Shukor, Z. Tasir, H. Van der Meijden, and J. Harun, "A predictive model to evaluate students' cognitive engagement in online learning," *Proc. Social Behav. Sci.*, vol. 116, pp. 4844–4853, Feb. 2014.
- [20] D. Yang, J. C. Richardson, B. F. French, and J. D. Lehman, "The development of a content analysis model for assessing students' cognitive learning in asynchronous online discussions," *Educ. Technol. Res. Develop.*, vol. 59, no. 1, pp. 43–70, Feb. 2011.
- [21] S. Masrom, A. Sani, and A. Rahman, "An adaptation of agent-based computer-assisted assessment into E-learning environment," *Int. J. Educ. Inf. Technol.*, vol. 3, no. 3, pp. 1–8, 2009.
- [22] *Bloom's Taxonomy Levels of Learning*. Accessed: Oct. 10, 2020. [Online]. Available: <https://kodosurvey.com/blog/blooms-taxonomy-levels-learning-complete-post>
- [23] Coursera: Machine Learning Lecture#9. *Neural Networks: Learning*. Accessed: Oct. 26, 2020. [Online]. Available: <https://billy-inn.wordpress.com/2014/08/05/coursera-machine-learning-lecture9-neural-networks-learning/>
- [24] J. J. Appleton, S. L. Christenson, D. Kim, and A. L. Reschly, "Measuring cognitive and psychological engagement: Validation of the student engagement instrument," *J. School Psychol.*, vol. 44, no. 5, pp. 427–445, Oct. 2006.
- [25] A. R. A. Conway, K. Kovacs, H. Hao, K. P. Rosales, and J.-P. Snijder, "Individual differences in attention and intelligence: A united cognitive/psychometric approach," *J. Intell.*, vol. 9, no. 3, p. 34, Jul. 2021.
- [26] S. N. Kew, S. Petsangsri, T. Ratanaolarn, and Z. Tasir, "Examining the motivation level of students in e-learning in higher education institution in Thailand: A case study," *Educ. Inf. Technol.*, vol. 23, no. 6, pp. 2947–2967, Nov. 2018.
- [27] M. Spüler, C. Walter, W. Rosenstiel, P. Gerjets, K. Moeller, and E. Klein, "EEG-based prediction of cognitive workload induced by arithmetic: A step towards online adaptation in numerical learning," *ZDM*, vol. 48, no. 3, pp. 267–278, Jun. 2016.
- [28] D. R. Krathwohl, "A revision of Bloom's taxonomy: An overview," *Theory Into Pract.*, vol. 41, no. 4, pp. 212–218, Nov. 2002.
- [29] C. D. Wysocki, *A Study of Cognitive Engagement in Online Learning*. Pullman, WA, USA: Washington State Univ., May 2017.
- [30] M. Mohler and R. Mihalcea, "Text-to-text semantic similarity for automatic short answer grading," in *Proc. 12th Conf. Eur. Chapter Assoc. Comput. Linguistics (EACL)*, 2009, pp. 567–575.
- [31] S. Shuhidan, M. Hamilton, and D. D'Souza, "A taxonomic study of novice programming summative assessment," in *Proc. 11th Australas. Conf. Comput. Educ.*, vol. 95, Jan. 2009, pp. 147–156.
- [32] C. Cabo, "Quantifying student progress through Bloom's taxonomy cognitive categories in computer programming courses," in *Proc. ASEE Annu. Conf. Expo.*, 2015, Art. no. 24632, doi: 10.18260/p.24632.
- [33] M. Farrus and M. R. Costa-jussa, "Automatic evaluation for E learning using latent semantic analysis: A use case," *Int. Rev. Res. Open Distrib. Learn.*, vol. 14, no. 1, pp. 239–254, 2013.



RIMSHA SHAHZAD received the B.S. degree from the Department of Computer Engineering, University of Engineering and Technology (UET), Lahore, Pakistan, in 2019, where she is currently pursuing the master's degree with the Department of Computer Science. Her research interests include machine learning, cognitive assessment, and text analysis.



MUHAMMAD ASLAM received the Ph.D. degree in computer sciences from CINVESTAV-IPN, Mexico, in 2005.

He has more than 15 years of experience in software architecture design, team leading, team building, and software project; and 14 years of experience in research and development and teaching at postgraduate level (supervising Ph.D. and M.Sc. thesis). He won the merit scholarship from the Board of Intermediate and Secondary Education, Sargodha Division, Pakistan, from 1984 to 1986. He was also received the Cultural Exchange Scholarship between Pakistan and Mexico, from 2000 to 2004, for Ph.D. studies. He already has supervised six Ph.D. and more than 50 master's thesis and final year projects. He also won the research grants from HEC and UET. He has published his research findings in number of well reputed impact factor journals of Springer, IEEE, and Elsevier. His research and teaching interests include artificial intelligence, distributed intelligence, knowledge-based systems, expert systems, intelligent agents, human-computer interaction, machine learning, computer-supported cooperative work, cooperative writing and authoring, cooperative learning, and distributed computing.

Dr. Aslam won Silver Medal from the Faculty of Agricultural Engineering, University of Agricultural, Faisalabad, Pakistan, from 1987 to 1991.

SHAHA AL-OTAIBI (Member, IEEE) received the M.S. degree in computer science and the Ph.D. degree in artificial intelligence from KSU. She is an Associate Professor with the Department of Information Systems, College of Computer and Information Sciences, PNU, Saudi Arabia. She also has SFHEA and a Senior Fellow Recognition from the U.K. Higher Education Academy. She is a reviewer in some journals and an editorial board member in other journals. Her main research interests include data science, artificial intelligence, machine learning, bio-inspired computing, cybersecurity, and information security.



SAEED ALI BAHAJ received the Ph.D. degree from Pune University, India, in 2006. He is an Associate Professor with MIS Department, COBA, Prince Sattam Bin Abdulaziz University, Saudi Arabia. His main research interests include information management, forecasting, information engineering, and information security.



MUHAMMAD SAQIB JAVED is a Lecturer with the Department of Computer Science and IT, Virtual University of Pakistan. He is also the Ph.D. Scholar with UET Lahore. He has 12 years of teaching experience at university level and mentoring undergraduate and graduate level students. He has supervised many master's level undergraduate projects along with M.S. level thesis. He has developed two graduate level courses (Introduction to Network Design and Management and Routing and Switching). He is a topper at M.Phil. level in June 2016 at NCBA & E Lahore by achieving 4.00 CGPA both in course level and thesis level. He has several publications in different journals. His research interests include multiagent systems, AI, machine learning, IoT systems, and cognitive networks.

TANZILA SABA (Senior Member, IEEE) received the Ph.D. degree in document information security and management from the Faculty of Computing, Universiti Teknologi Malaysia (UTM), Malaysia, in 2012. She is currently a Full Professor with the College of Computer and Information Sciences, Prince Sultan University (PSU), Riyadh, Saudi Arabia, and also a Leader of the AIDA Laboratory. She has published over 300 publications in high-ranked journals. Her primary research interests include bioinformatics and data mining and classification using AI models. She received the Best Student Award from the Faculty of Computing, UTM, in 2012; and the Best Researcher Award from PSU, from 2013 to 2016. She is an editor of several reputed journals and on a panel of TPC of international conferences.

...



AMJAD REHMAN KHAN (Senior Member, IEEE) received the Ph.D. degree, specializing in information security using image processing techniques from the Faculty of Computing, Universiti Teknologi Malaysia (UTM), Malaysia, in 2010. He is currently an Associate Professor with CCIS, Prince Sultan University, Riyadh, Saudi Arabia. He is also a PI in several projects and completed projects funded by MoHE, Malaysia, Saudi Arabia. His research interests include bioinformatics, the IoT, information security, and pattern recognition. He received the Rector Award for the 2010 Best Student from UTM.