International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Analysis of Bloom Taxonomy-Based Examination Data Using Data Mining

[1]Amit Kumar, [2]Dr. Dinesh Singh, [3]Dr. M. S. Dhankhar

Submitted: 15/09/2023 **Revised**: 22/10/2023 **Accepted**: 08/11/2023

Abstract: Question classification means, the selection of a category of questions from a list of established question categories. It is a unique kind of text categorization in which there are considerable differences between the two forms, especially when the test questions comprise only a few terms that include or express the substance of the question. Consequently, creating exam questions is a stage that academics find very difficult. Thus, Bloom's Taxonomy has become a framework for creating examinations that span a broad variety of cognitive levels based on the various abilities of students. Using data mining, this research proposes a strategy making use of Bloom's Taxonomy to categorise exam questions cognitive levels. In this work, a CNN-model-based BiLSTM classifier is utilized to categorize questions using feature selection techniques. By employing Mutual Information feature selection, the BiLSTM classifier obtained the best classification model performance using the macro F1-measure. In conclusion, this study's trials demonstrate that the feature selection approaches contributed positively to the performance of the classifier.

Keywords: *BiLSTM, CNN, Bloom's taxonomy, Feature extraction, data mining.*

1. Introduction

There are several methods of evaluation used to determine the learning progression. The written exam is the most prevalent kind of evaluation and plays a significant role in evaluating the semester-long cognitive performance of students [1] [2]. Extensive use of Bloom's Taxonomy layout examples is seen in education [3], [4]; it facilitates the creation of computer science-related questions and enhances curriculum design and assessment. The three domains that comprise Bloom's Taxonomy are Cognitive, Affective, and Psychomotor Learning. Determining a question's level typically involves manually calculating the Bloom's Taxonomy's Cognitive Complexity.

Manual categorising of questions involves a substantial amount of the era of big data sets. Moreover, perception differences in categorising the inquiries led to a manually variable categorization procedure. In order to address these concerns, categorization of automation can be performed using NLP [5, 6]. The aim of this investigation is to assess the effectiveness of categorising query predicated on Cognitive Bloom's Taxonomy Level utilising the bidirectional LSTM algorithm with CNN for feature extraction. In addition, as a lecturer's approach for creating questions that would serve as a standard for measuring students' comprehension of the presented content based on learning goals [7], researchers provide the following: In this investigation, the dataset consists of questions from midterm and final exams given by instructors. These questions will be manually classified before being subjected to a series of pre-processing techniques, containing stemming, tokenization, filtering, and feature extraction. The existing data must then be converted into vector features, which are numerical data and then inputted into the bidirectional LSTM algorithm building model [8-9].

[1]*Research Scholar, Department of Computer Science and Engineering, DCRUST, Murthal, 19001901901amit@dcrustm.org*
[2]*Associate Professor, Department of Computer Science and Engineering, DCRUST, Murthal, dineshsingh.cse@dcrustm.org*
[3]*COE, DCRUST, Murthal, Coe@dcrustm.org*

This study's objective is to evaluate the existing literature on data mining techniques applied to the assessment of student test scores. The secondary objective of this endeavour is to create a data mining strategy for a testing system utilising Bloom's taxonomy. Here, the author attempts to implement the proposed strategy by using the right programmes and tools to improve the project's outcomes. To conduct research, academics recommended that a policy decision be made to construct the examination based on multiple levels, such as simple and tough. The study is conducted using BiLSTM and CNN model as a proof of concept; this ensures a better convergence rate than conventional models, to train the optimised deep learning with CNN model with positive and negative instances that evaluate the Rule mining approach, comparing the suggested model's performance metrics to those of other cutting-edge models, such as prediction accuracy and specificity. Keeping these objectives in mind, the proof-of-concept analyses data sets to demonstrate that Bloom taxonomy is applicable to the given scenario [10]. If this is the case, the use of Bloom taxonomy to predict pupil achievement can be studied in greater depth.

The remaining sections of the paper are arranged as follows: The research's literature review portion is included in the second section. The third section explains the research techniques and BiLSTM model with CNN feature extraction. The conclusions of the investigation are described in the fourth section titled "Results". The fifth section concludes the assignment and identifies potential research areas.

2. Literature Review

Researchers take a broad view of the research area's context and provide background information. In this section, researchers would examine the research that has been done on the topic of Data Mining of Examination Evaluation settled on Bloom's Taxonomy using several distinct Data Mining approaches. Moreover, this chapter provides an analysis of multiple research studies for Bloom's Taxonomy, Convolutional Neural Networks (CNN), and Naive Bayes (NB) that were written by a variety of authors.

Kamlasi et al. (2018) [11] discussed the study analysed the English examination questions according to the updated Bloom's taxonomy, as was explained. Qualitative descriptive research describes the investigation ideally. qualitative descriptive research. As a research instrument, the end-of-year English section secondary education exam was utilised. Coding, categorising, analysing, and deliberating were used by the scientists to examine the data. Verbs rather than nouns were chosen because Bloom taxonomy is intended to be implemented rather than solely described. 22 elements, or 44% of the total, were in the memorization taxonomy. Two tests (4%) were administered to gauge taxonomy comprehension. 21 entries, or 42% of the total, were generated using taxonomy. The analysis of taxonomy generated five elements. Even if taxonomy research and development produced no outcomes.

Laddha et al. (2021) [12] stated that acquiring an education is the initial stage towards knowing the truth and developing an analytical mindset. Cognitive process and knowledge dimensions of Revised Bloom's Taxonomy serve to classify the learning process into six distinct cognitive processes and four distinct knowledge dimensions, with an emphasis on the human capacity to reason. The objective of this study is to establish a comparative analysis of the categorization of the summative assessment based on Revised Bloom's Taxonomy using Convolutional Neural Networks, Long Short-Term Memory, and Deep Learning techniques.

Sami et al. (2020) [13] emphasised that the revised Bloom's Taxonomy is a framework that facilitates the development of learning objectives and is recommended for use in the field of technology by academic institutions worldwide. The theory establishes a hierarchy of educational purposes, which includes cognitive, sensory, and affective domains that enhance cognitive capacity and skill disparities among students. No research has been conducted on analysing students' RBT skills using data mining. Using descriptive and predictive data mining, the article evaluates each student's RBT competence. A classifier that breaks down question complexity into six categories—comprehending, applying, remembering, applying, assessing, analysing, and generating—is used to categorise students. The proposed classifier can identify RBT question levels with 98% precision.

Kumara et al. (2019) [14] discussed that examinations help towards instruction and learning. The knowledge and skills of students are evaluated via queries. Exam papers are frequently used by instructors in higher education to evaluate retention and application of students. Questions of a higher, intermediate, and lower order can be used to evaluate the cognitive abilities of young children. Based on Bloom's Taxonomy, the goal of research is to create a method for classifying test papers that is practical. For this investigation, information technology literature was analysed. The study's goal was to ascertain if the assessment questions were correctly categorised using Bloom's Taxonomy.

Ullah et al. (2019) [15] acknowledged that instructors encounter challenges when attempting to evaluate the computer programming skills of their students, particularly at the introductory level where student enrolment is frequently high. In light of Bloom's taxonomy, the research suggests a different approach for evaluating students' programming skills. The presented criteria serve as the foundation for the originality of the delivered technique. This is due to the fact that they are founded according to Bloom's cognitive level assessment, the learned skills are graded on a scale. This study employed a rule-based assessment technique that mapped students' proficiency to the appropriate cognitive levels based on the code, using an automated decision-making procedure. As opposed to earlier research that made use of cognitive levels as a framework for question creation but still relied on human assessors to determine whether or not a student was competent, this study utilised artificial intelligence to determine whether or not a student was competent.

Ali et al. (2019) [16] concluded that the general concept of question classification is choosing a question category from a predefined list of categories. Due to the various ways in which such a text classification differs from traditional text classification, test questions consisting of only a few sentences that convey or embody the question's content are particularly interesting. In this work, machine learning researchers provide a mechanism for automatically categorising questions into Bloom's Taxonomy areas. To divide the questions into groups, Nave Bayes and K-Nearest Neighbour classifiers are employed together with Chi-Square, Mutual Information, and Odds Ratio feature selection algorithms. Feature selection methods improved classifier performance as expected at the beginning of the investigation.

Zahid et al. (2020) [17] the majority of computer science students, particularly those in their first year, lack programming skills, according to study. Most pupils lack basic programming skills when they first attend the classroom, and they are unfamiliar with the new environment that comes with creating programmes in a language that is syntax-specific. The annual rate of failure is influenced by multiple characteristics. Thus, computing programming education is evaluated and enhanced by the application of Bloom's taxonomy. The paper presents a fresh way to programming assessment methods by directly mapping a learner's competency level to the appropriate cognitive stages of Bloom's taxonomy, without the need for a priori question mapping. These findings show that Bloom's taxonomy is an effective teaching and evaluation tool for programming.

Prasad et al. (2021) [18] to asses a taxonomy called Bloom's was created to assess a student's intelligence. For the task to be completed on time, automation is needed. Language written by humans is automatically processed by a branch of computer science called Natural Text Processing, or NLP. To name a few applications, there is sentiment analysis, fake news detection, language translation, and grammatical error detection. Text that is too big to handle is divided into smaller bits using a technique called tokenization. Through the use of Bloom's cognitive level, question items were categorised in the investigations. 70 questions make up the training package exam. Question types are categorised based on standards using Taxonomy verbs. For evaluating the cognitive complexity of the test, the research provides a strategy. Trainers can use it to learn about the cognitive level of their students.

Makhlouf et al. (2020) [19] stated that one of the most popular frameworks for developing and accessing educational exams is still Bloom's Taxonomy (BT). Recently, a large amount of time and effort have been expended by several academics mechanising the procedure for classifying test queries into groups according to BT. A comprehensive overview of researchers is included in the paper. Some of the techniques commonly employed in QC into BT cognitive levels are preprocessing, dimensionality reduction, categorization, and assessment. The integration of multiple methodologies can lead to an improvement in quality control accuracy. The goal of future research is to expand the instruction set by experimenting with various question types. Combining multiple feature extraction/selection techniques is advised.

Atachiants et al. (2016) [20] claimed that optimising software for multicore computers is becoming more and more challenging for programmers. Performance optimisation and debugging parallel applications are challenging. Tools for simultaneous application development allow multicore programmers to identify and categorise performance problems. Our analysis shows that there is consensus among experts and developers on how to identify parallel performance issues. The literature review summary is shown in Table 1.

Tanalola et al. (2017) [21] discussed that the research looks at how well writing successful test questions is guided by Bloom's Taxonomy. A good and relevant exam would provide many levels of challenge-to-challenge pupils of varied abilities. Here, the exam's complexity is determined by the keyword(s) included in each question. Information and keywords would be mined from the textual content of the exam paper using a knowledge-based technique and text-mining technology. The instructor can evaluate the test questions and create a comprehensive exam using Bloom's Taxonomy.

Uma et al. (2017) [22] stated that the success of a school and its pupils can be measured in large part by the quality of its teaching and learning. A significant overhaul of an assessment system is required to meet the high expectations put on instructors and to reduce disparity between the method of teaching and learning and the results of that process. Assessment, being an important indication of achievement, should provide students an opportunity to use their critical thinking skills. In this study, researchers take a closer look at the critical thinking skills measured using a revamped version of Bloom's Taxonomy. In addition, the evaluation's complexity is analysed to establish the system's overall quality. In this study, researchers use a weighted data mining approach to systematically classify Bloom's taxonomies and associated logical apex states.

Bindra et al. (2017) [23] stated that one of the automated architectural designs in this research finds the right response to the test questions based on the student's current grade. Direct assessment of student learning relies heavily on the whole examination process. Therefore, it is relatively necessary to prepare a thorough test paper and its arrangement. Predicting student ability, education system growth, improved teaching methods, future student interest, etc. might all benefit from this data. OneR, ZeroR, j48, Naive Bayes, and Instance-Based Learner (IBk) are only a few of the data mining classifiers used to forecast the final grade in the course. Time, accuracy in detection, and classification error are only a few of the metrics taken into account in the comparative study.

Jayakodi et al. (2016) [24] discussed that assessment is often regarded as the most significant indicator of students' actual performance in the classroom. Questions like this are the most popular kind of assessment in schools. The examiners have the challenging challenge of coming up with test questions that would achieve the desired learning goals for the course. Consequently, principal objective of the research is to automatically classify test questions according to their learning levels using Bloom's taxonomy. Before developing the rule set for this classification, Natural Language Processing (NLP) methods like stop word removal, tokenization, tagging, and lemmatization were used. WordNet similarity techniques using Natural Language Toolkit (NLTK) and the cosine sameness formula were designed to provide a specific set of criteria for identifying each exam question's category and weight according to Bloom's taxonomy. Over 70% accuracy was reported by the developed rule set.

Patil et al. (2017) [25] discussed that learning new facts is important, but so is learning how to analyze and evaluate those facts. Learning is broken down into six levels, or "levels of cognitive competence," Revisions to Bloom's Taxonomy, with each level reflecting a more complex level of cognitive complexity. These levels are arranged in ascending order of difficulty. Lorin Anderson and David Krathwohl are the writers of the most recent version of Bloom's Taxonomy. It provides a scoring system for assessing an individual's own cognitive and knowledge abilities. The main purpose of this research is to assess how well the SVM and K-NN machine learning approaches classify a question bank using Revised Bloom's Taxonomy, as well as how accurate and efficient. Here table 1 shows the summary of the literature review.

Table 1: A Summary of the literature review

Writer	Techniques	Outcomes
Kamlasi et al. (2018) [11]	BT	The taxonomy of memorization consists of 44 entries. It was tried taxonomy twice (4%). Out of 21 items (42%), taxonomy produced 21. 5 entries were found using taxonomy.
Laddha et al. (2021) [12]	DLM, CNN, LSTM	The primary purpose of this research is to compare the Revised Bloom's Taxonomy-based classification utilised in the summative assessment..
Sami et al. (2020) [13]	Revised Bloom's Taxonomy (RBT)	The suggested classifier has a 98% accuracy rate in identifying RBT input question levels.
Kumara et al. (2019) [14]	RBT (Revised Bloom's Taxonomy)	The purpose of the study was to ascertain whether exam questions were correctly classified using Bloom's Taxonomy.
Ullah et al. (2019) [15]	Rule-based testing, or RBT	Based on the study's rule-based evaluation technique, an autonomous decision-making system converts students' skill levels into corresponding cognitive levels in the written code.
Ali et al. (2019) [16]	NB, KNN classifiers are utilised, along with Chi-Square, Mutual Information, and Odds Ratio.	Researchers present a machine learning-based method for automatically classifying queries into Bloom's Taxonomy domains.

Zahid et al. (2020) [17]	RBT	As demonstrated by the results, Bloom's taxonomy is a useful tool for programming assessment and instruction.
Prasad et al. (2021) [18]	NLP	Using Taxonomy verbs, the method groups questions based on the standards.
Makhlouf et al. (2020) [19]	BT	By using BT techniques, quality control accuracy can be increased.
Atachiants et al. (2016) [20]	RBT, NLP	Our analysis shows that there is consensus among experts and developers on how to identify parallel performance issues.
Tanalola et al. (2017) [21]	Bloom's Taxonomy	The instructor can evaluate the test questions and create a comprehensive exam using Bloom's Taxonomy.
Uma et al. (2017) [22]	RBT	It assesses the degree of complexity found in an evaluation and establishes the system's quality. In this work, Bloom's categories and associated thinking levels are categorised using a weighted data mining technique.
Bindra et al. (2017) [23]	automated architectural designs	Time, accuracy in detection, and classification error are only a few of the metrics taken into account in the comparative study.
Jayakodi et al. (2016) [24]	NLP, NLTK	Over 70% accuracy was reported by the developed rule set.
Patil et al. (2017) [25]	RBT, SVM, KNN	Over 80% accuracy was reported by the use of SVM, KNN, and RBT.

2.

3. Research Methods

3.1 Data mining

The term "data mining" refers to a group of methods that can be applied to very large and difficult datasets. This is done to eliminate randomness and reveal the underlying pattern that had been hidden. These methods of data mining often need extensive amounts of processing power [26].

To discover patterns in the data, researchers employ various technologies, techniques, and theories related to data mining. There are an excessive number of things at play here. This is one of the primary reasons why data mining has developed into such an important academic discipline.

3.1.1 Data mining includes the KDD process model for knowledge finding in Databases.

The study written by Fayyad et al. [26] served as the foundation for the KD process models that came after it. The Knowledge Discovery in Databases (KDD) model that it devised did not make any specific reference to a data mining (DM) methodology, but it did make the knowledge creation process simpler and more iterative [27]. The KDD model can be broken down into five distinct stages, as represents in Figure 1, data transformation, data processing, data mining, data sampling and selection, and assessment. During the data selection phase, an analyst would search through a sizable data bank and choose pieces of information from there that are pertinent to the knowledge discovery process. Dealing with noisy and missing data is one of the tasks involved in the data processing. This task also ensures that the appropriate input is employed throughout the KDD technique, which results in the production of valid output. The phase of data transformation is when methods like dimension reduction and transformation come into play. These approaches are used to figure out which characteristics are the most useful. Data mining is the process of choosing a data mining task that fits the analytical objectives, choosing an algorithm or methods with appropriate parameters, and using these tactics to draw patterns out of the data. The final step of the evaluation process is to analyse the collected patterns and derive useful information from them; subsequent rounds can include corrective actions.

The KDD approach emphasizes a data analysis process that is interactive and iterative activities and is essentially data-centric. However, although data-related responsibilities are defined, this sketch model lacks a business viewpoint [28]. In the comprehensive version of the model, both pre- and post-processing procedures are included. In the first phase, the focus is on acquiring knowledge of the application domain and selecting project objectives from the customer's perspective. In the last stage, the focus is on this consolidating the acquired intelligence, which should be incorporated into the relevant systems and documented.

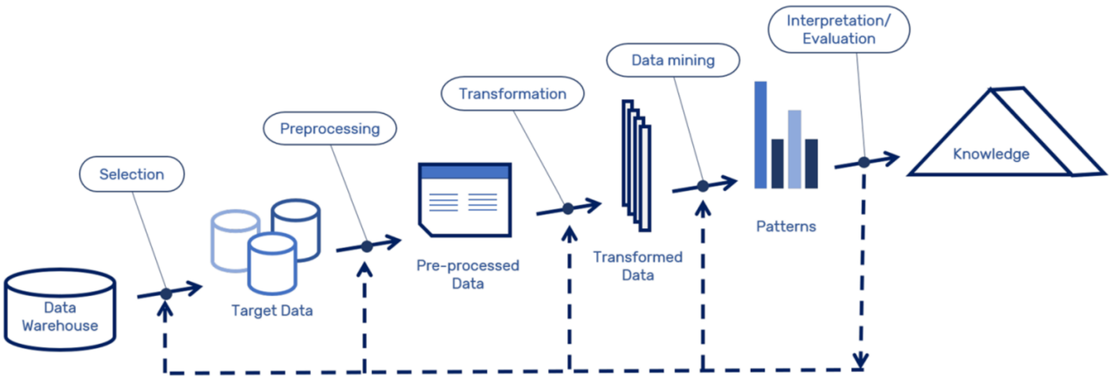

Fig 1: KDD Process model [27]

The technique for data mining consists of several phases. Gathering raw data is supposed to result in the acquisition of some new kind of knowledge as the desired outcome of this process [27]. The iterative process can be broken down into the following stages:

- **Data cleaning**: The information gathered during this procedure stage will be filter to remove extraneous and noisy information.
- **Data Integration**: The integration of several data sets takes place in the same area.
- **Data Selection**: The data that is going to be used for the analysis have to be picked out of the collection of data and then extracted.
- **Data Transformation:** In addition to that, it is a method for consolidating data. During this stage, the selected data are transformed into various forms. That is an appropriate approach for the mining procedure [29].
- **Data Mining:** To extract patterns that could be possibly significant, researchers need to apply intelligent techniques.
- **Pattern Evaluation:** The approach locates knowledge-representing patterns that are fascinating by using the metrics that are supplied as the foundation for the search.
- **Knowledge Representation:** It is the last stage. During this phase, in particular, information is found and given to the user. The key stage involves visualization approaches. That aids users in understanding and interpreting the findings of data mining.

3.2 Bloom's Taxonomy

Bloom's Taxonomy is a taxonomy that categorises a wide range of knowledge, including that which is learned, considered, and understood. Professionals incorporate Bloom's taxonomy is incorporated into lesson plans, curricula, and evaluation protocols to service students more effectively. [30].

"Ben Bloom, Ed Furst, Max Englehart, David Krathwohl, and Walter Hill" constructed the first edition of Bloom's Taxonomy in 1956. This concept, or more accurately the educational model, organises the skills and knowledge that must be obtained prior to acquiring new information.

3.2.1 Bloom's Taxonomy Three Domains

- ✓ **Cognitive Domain:** At the cognitive states of Bloom's taxonomy, the focus is on learning and enhancing intellectual ability. There are six distinct classifications of cognitive complexity [31]. The Cognitive Domain is shown in Figure 2.
- ✓ **Knowledge:** Acquiring specific knowledge like facts, statistics, and essential concepts.
- ✓ **Comprehension:** Understanding the newly acquisition of information at the stage of acquiring knowledge.
- ✓ **Utilisation:** Utilizing one's abilities, ideas and principle most productively and fruitfully is technically feasible.
- ✓ **Analysis:** Conducting in-depth research into the programmed, drawing specific judgements about it, and learning its many interconnected parts.
- ✓ **Evaluation:** Formulating conclusions and providing justifications based on program-generated information, and using that information to shape those judgments and justifications.
- ✓ **Creation:** The process of designing, developing, and constructing the actual application in use that generates novel outcomes.

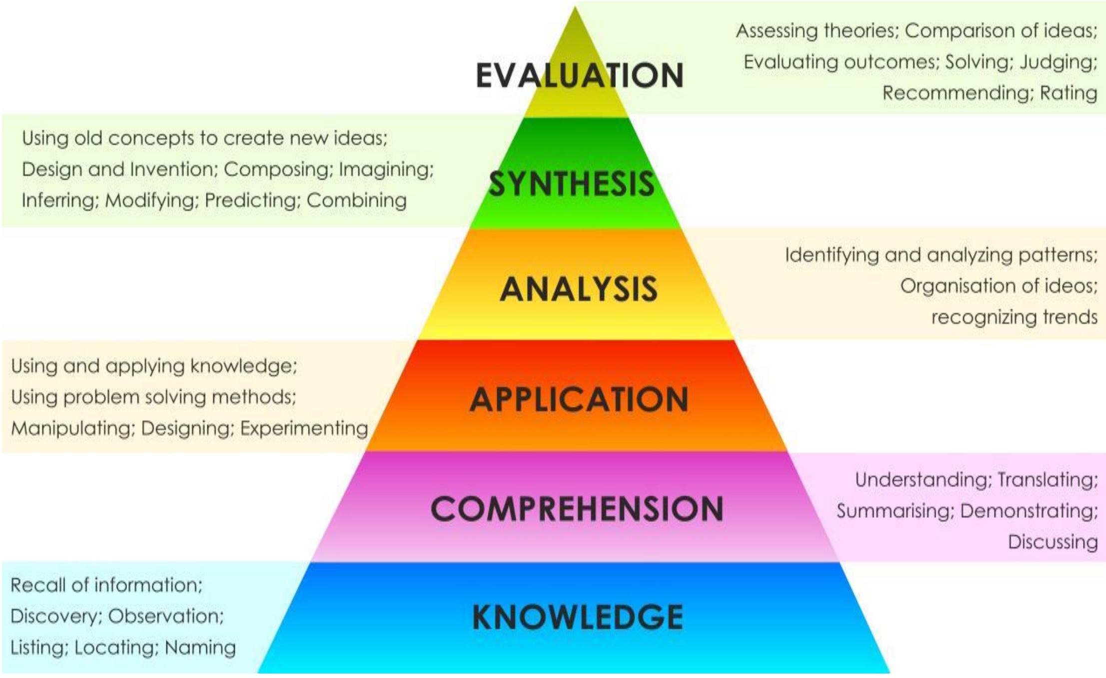

Figure 2: Cognitive Domain [31]

- **Affective Domain**

At these states of Bloom's taxonomy, the emphasis shifts to the reader's individual responses to the text. It presents individuals, events, and perspectives that can be interpreted in numerous ways. The emotional spectrum is comprised of the following essential characteristics. [31] Figure 3 depicts the Affective Domain.

The five stages, from lowest to highest with examples, are as follows:

- Receiving: On the first day of school, paying attention and making an attempt to learn the names of the new classmates are instances of the "basic awareness" we're referring to.
- Responding: Participated actively and reacted to stimuli with an emphasis on responding, for example: participating in a classroom discussion.
- Valuing: The significance of a specific object or item of information, which can range from basic acceptance to complex commitment; values are inextricably linked to prior knowledge and experience. For instance, embracing variety and respecting the backgrounds and perspectives of others.
- Organizing: Accepting professional ethical standards is an example of arranging values in ascending or decreasing order of significance and establishing an original value system with an

emphasis on opposing and linking previously established values.

- Characterizing: The fifth and final layer consists of combining the information from the previous four levels into an abstract body of knowledge; at this point, the value of system is fully operational and influences the system's behavior. This can be demonstrated by adhering to a code of ethics while on the job.

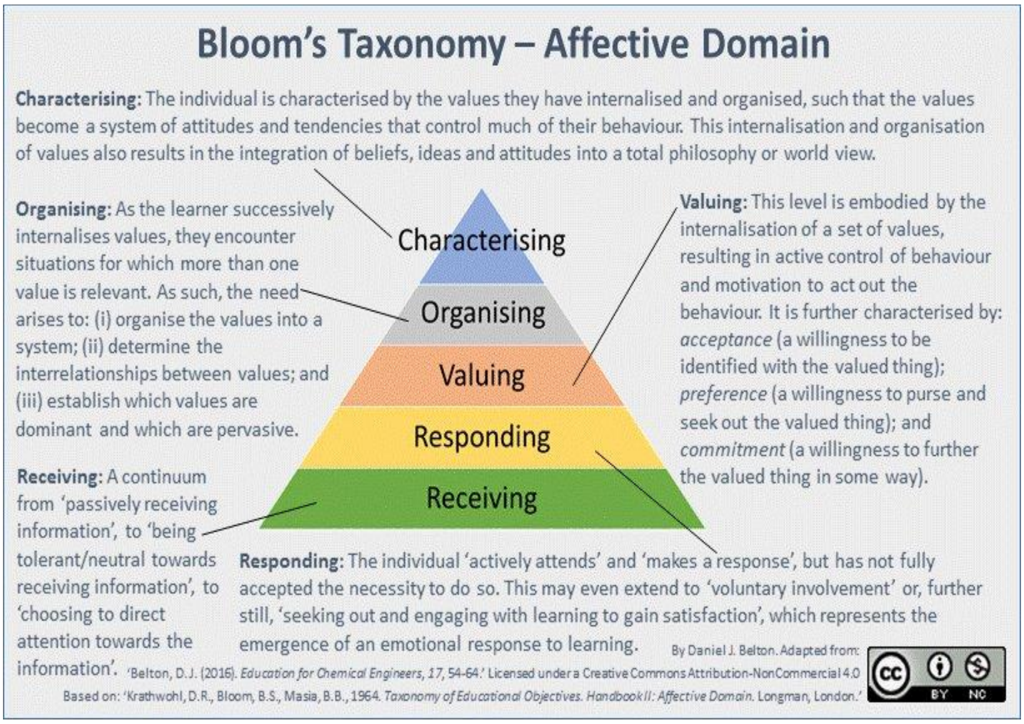

Figure 3: Affective Domain [31]

- **Psychomotor Scales/Domain**

Coordination, sensory organ motion, and the student's own physical movement are all part of the psychomotor domain of Bloom's taxonomy. Researchers would have to put in a lot of time and effort before the author could become adept with these attributes. Using a computer, playing an instrument, and driving a car are examples of psychomotor skills. The psychomotor model, like the cognitive model, is not devoid of modifications. This paradigm, whose beginnings date back to 1970 when it was published by Robert Armstrong and colleagues, includes five stages, and Figure 4 illustrates all five stages of the Psychomotor Domain [32].

The following is a list of the seven levels, along with some instances of each:

- Perception: Knowledge of one's environment; for example, predicting where a thrown ball would land and altering one's actions accordingly to be in a position to catch it.
- Set: Mental, physical, and emotional dispositions that set you up to function as you do. One such goal is the realization that it is currently unattainable, such as the desire to learn how to throw a perfect strike.
- Guided Response: The initial phases of acquiring a physical skill. One can only learn by making mistakes. Consider the process of learning how to throw a ball by observing a coach and paying close attention to the form.
- Mechanism: The level of proficiency that sits between novice and expert. It requires transforming learned responses into reflexes so that the current task can be performed with confidence and skill. For example, the ability to throw a baseball to the catcher without missing.

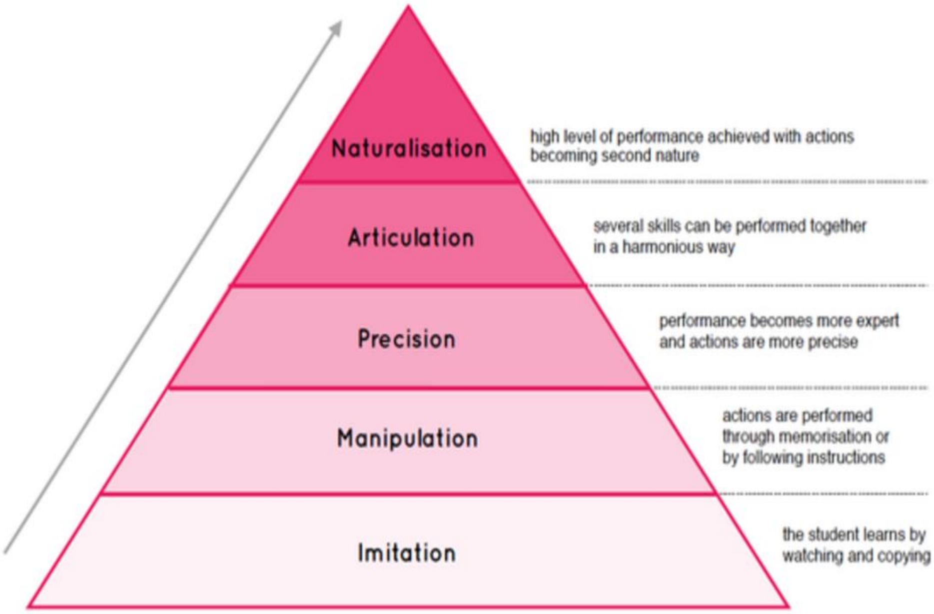

Fig 4: Psychomotor Domain [32]

- Complex Overt Response: Automatically and without thought performing tough tasks, like delivering a superb strike right into the catcher's mitt.
- Adaptation: A talent is considered to be "modifiable" when it has been honed to such a degree that it can be changed to fulfill certain requirements, for instance. This would include the ability to throw a perfect strike to the catcher even with a batter at the plate.
- Origination: the ability to develop new movements in response to a challenge or condition. These movements are derived from a set of previously taught bodily movements. For example, learning to throw the perfect curveball involves the same skill set as learning to throw the ideal fastball [32].

3.2.2 Bloom’s Taxonomy's Six Categories

Consider a situation in which an institution's student is assigned an essay to evaluate their competency in the art of communication. To effectively complete the assignment, the student is expected to know the following levels of Bloom's taxonomy [33]:

- ❖ **Recall:** Make use of your prior knowledge and the many forms of communication you will require on a regular basis.
- ❖ **Sympathize:** Consider the various interaction approaches that users can employ to attain the goal.
- ❖ **Utilize:** Utilize a credible theory of communication supported by scholarly research, and put it to the test in terms of how it typically communicates with others.
- ❖ **Examine:** Consider the fact that, depending on the circumstances, the target demography, and the objective, it can be essential to employ a variety of communication channels to accomplish the desired result.
- ❖ **Evaluate:** To grasp the problems more thoroughly that are presently being addressed, it can be useful to take a step back and examine the communication process in its entirety.
- ❖ **Construct:** Construct a model of one's mode of expression and utilise it as a guide.

3.3 Bi-directional LSTM

BiLSTM (Bidirectional “LSTM) is a recurrent neural network that is mostly used for natural language processing. Unlike regular LSTM, input travels in both ways, and information from both sides can be used. Additionally, it is a potent instrument for modelling the sequential dependencies between words and sentences in both directions.

In short, BiLSTM adds an LSTM layer that reverses the flow of information. In short, it indicates that the input sequence flows in reverse in the extra LSTM layer. The outputs from both LSTM layers are then combined in a variety of methods, such as by averaging, summing, multiplying, or concatenating.”

As an example, BiLSTM 's shown in Figure 5 below:

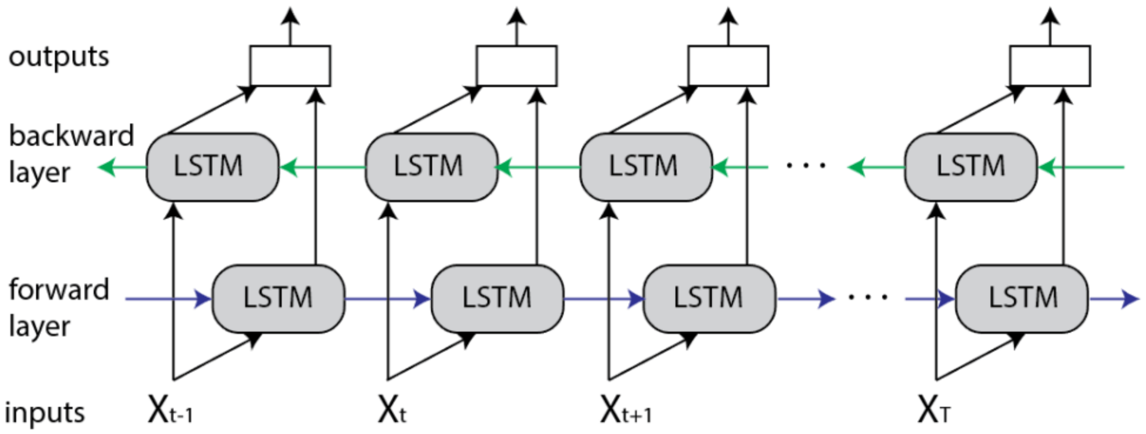

Fig 5: BILSTM model architecture [34]

3.4 Research Methodology

To illustrate, see Fig. 6. Which provides a comprehensive description of the recommended course of action. Following the receipt of the examination datasets, the author first opens the test papers, and then the text files containing the questions. The development of questions comes first in the process of putting the suggested model into action, which is then followed by the extraction of keywords. Tokenization, the elimination of white space, the removal of punctuation marks, the removal of stop words, and the deletion of non-letter characters were all conducted during the process of preparing the question.

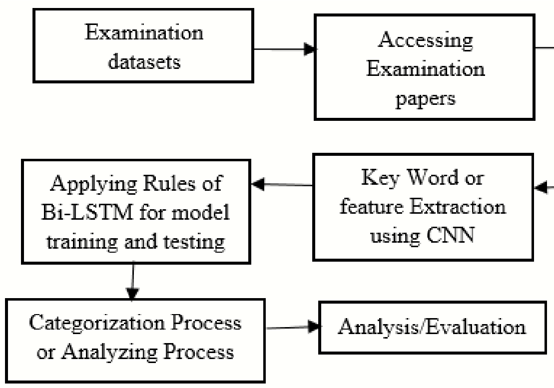

Fig 6: Proposed model for question classification

These basic guiding criteria were laid forth for the first time. The researchers drew on their prior expertise and understanding of various cognitive levels to formulate guidelines for responding to lower-order queries. A rule for intermediate-order inquiries was developed with analytical and application-level keywords serving as the basis for the development of the rule. The rule for the third level of questions was developed by drawing on concepts from both the first and second levels of questioning. This activity is carried out as a component of the process of applying the regulations. One of the components of the criteria is represented by each of the groups. After then, the questions are graded according to how challenging an examiner would likely find them. Exam fairness was established by looking at the proportion of questions that fell into each of two categories: simple and demanding. The reason behind this choice is as follows and the following diagram illustrates the procedure for the proposed model:

- **Questions Pre-processing**

The dataset consists of many questions and these questions are having different levels are easy, moderate, and challenging levels of difficulty. So now the dataset is going to send to the pre-processing unit.

In the data pre-processing step, the database is accessed by three modules, which are as follows:

- Removal of duplicate records: If any question record seems to be a duplicate of another record, the pre-processing unit would remove the duplicate record from the database and retain the original record.
- Removal of unwanted symbols: If any symbols are found in the record of the question, the pre-

processing unit would remove them and retain the original record in the database. Here symbols imply are @#$ %^&*()":><? Etc.

- Removal of URL: URL removal is the removal of a URL from inside any text if any URL is present as a hyperlink way.

In preparation for data mining, issues with missing, noisy, or otherwise inconsistent data are resolved during pre-processing. There are two components to the text pre-processing phase. These processes are referred to as "cleaning" and "tokenization" of text.

Exam papers can contain numerous graphs, charts, and equations. This investigation centred on textual questions. Text files containing unnecessary information, such as tables and figures, need a text-cleaning procedure. The entire project was completed by hand. As part of text preparation, tokenization divides lengthy text strings into tokens that are more manageable. The eventual goal of tokenization is word exploration within a sentence. It is feasible to tokenize a paragraph into sentences and then each sentence into words. In most instances, additional processing can be performed on a text once it has been properly tokenized. Researchers eliminated interword gaps, punctuation, numerals, whitespace, and non-letter characters as part of the normalization procedure.

- **Keyword or Features Extraction**

In 1950, Benjamin Bloom [27] created a system for categorizing learning objectives based on the level of student comprehension. This organizational chart consists of six levels. Numerous sorts of researchers, educators, curriculum designers, and evaluators have found this method useful Intelligence, Comprehension, and Application represent the initial three levels of learning LOTS. The final three Higher Order Thinking Skills levels are titled Analysis, Synthesis, and Evaluation. Educators must be familiar with conventional classification systems such as Bloom's Taxonomy to provide a healthy range of difficulty in examinations. The each question's level of difficulty was determined by compiling a list of exam paper keywords. For extracting keywords or features for the further process here researchers used CNN for keyword extraction and with the help of CNN, researchers can able to extract those features or keywords that would help in training the BILSTM model.

- **Rules Development in BiLSTM model and analyzing process**

The algorithms BILSTM were designed to follow Bloom's Taxonomy [27] criteria for all six levels. In this study, six new recommendations were created based on the keywords identified inside the six states of Bloom's Taxonomy. The mentioned grammatical rules enhance and simplify the outputs of algorithmic categorization. Bloom's Taxonomy was used to analyze its grammatical structure so that rules could be created. The rules would define which terms to utilize for each search. The development of rules was based on syntactic patterns.

A question could include certain significant phrases (VB: Verb Phrase (basic form), DT: Determiner lingua franca), NN: Noun Phrase (single or plural), JJ: Adjective Phrase, ADV: Adverb Phrase, PP: Prepositional Phrase. Phrase). A query might break into an individual phrase to develop the laws. The rules were created based on the phrases. Here BiLSTM model is used for the classification of the questions and this model is considered all six states or levels of bloom's taxonomy as keywords for classifying the queries. It means with the help of these six levels BiLSTM model can easily demonstrate how many questions are in levels hard, medium, and easy. The taxonomy of Bloom's six states are as follows:

a) Synthesis Level

b) Analysis Level

c) Application Level

d) Knowledge Level

e) Evaluation Level

f) Comprehension Level

Following the establishment of six rules, these rules were consolidated into three fundamental rules or parameters based on the findings of the proposed model; Higher Order Questions (HOQ) as hard-level questions, Intermediate Order Questions (IOQ) as medium-level questions, and Lower Order Questions (LOQ) as easy level questions were the three kinds of questions provided. BiLSTM model has been allocated Bloom's Taxonomy has six cognitive levels for analysing queries in these three core rule areas of research. Questions of a Lower Order predict by Comprehension and Knowledge Levels. Questions of Intermediate and Higher Order predict by Application, Analysis, Synthesis, and Evaluation levels. So, in this way rules are assigned for predicting the different levels of questions: Rule 1 and Rule 2 were utilized as one rule for Lower Order Questions. Rule 3 and Rule 4 were combined into a single rule for IOQ. Rule 5 and Rule 6 were combined into a single rule for HOQ.

- **Determine the Question Paper's State (Evaluation State)**

In this stage, after the completion of the process of the author's project, here the author would be able to tell how the level of the question is. This means the level of questions is easy or medium or hard with the help of different results and graphs.

4. Dataset

Table 2 depict the dataset which is described the types of question difficulty in terms of easy, medium, and hard. The questions are compiled from a variety of academic institutions' examinations.

Table 2. Dataset

S. No	Types of question difficulty
1	Easy
2	Medium question difficulty
3	Hard

5. Results and Discussion

The questions level prediction model was performed using the main model as the BiLSTM model, CNN algorithms for feature extraction, and BiLSTM model for model training and testing, here the findings are presented below for author's perusal. Within the application, the training data served the purpose of developing a model, which was subsequently utilized to analyze the test data. These results were compared to those that were produced by the machine learning algorithms included in the program known as BiLSTM and LSTM models algorithms. When comparing the effectiveness of several machine learning algorithms, standardized parameter values were used.

Table 2: The comparison criterion applied to the outcomes.

CRITERIA	MATHEMATICAL EQUATION
Number of Correctly Labelled Data / Total Number of Data	$(TP+TN)/(TP+FP+TN+FN)$
Accuracy Percentage	$((TP+TN)/(TP+FP+TN+FN))\times 100$
Accuracy(Precision)	TP/(TP+FP)
Recall	TP/(TP+FN)
F-Measure	(2×TP)/(2×TP+FP+FN)

The findings were evaluated using the most prominent standards in academic research and the overarching objective of suggesting several computer models and evaluating them to predict questions level. Table 2 outlines these criteria and computations for your convenience. Various outcomes highlight the disparity between the results according to the different parameters such as easy-level questions, hard-level questions, and medium-level questions. Additionally, "further findings are demonstrating the effectiveness of the proposed model with respect to f1 score, accuracy, precision, recall, and confusion matrix."

The below results illustrate the discrepancy between the results according to the different parameters.

The research's illustrated outcomes highlight the various parameters that were used to gather data. So, the first result in figure 7 is showing overall precision of the suggested model with 98% of training precision and 82% of validation accuracy. Here by this, the author is trying to show that our proposed model outperforms the base model's accuracy. As same as the second result in figure 8 is showing the overall loss percentage of the proposed model with 0% of training loss and 20% of validation loss. Here by this, the author is trying to show that our proposed model outperforms the base model loss.

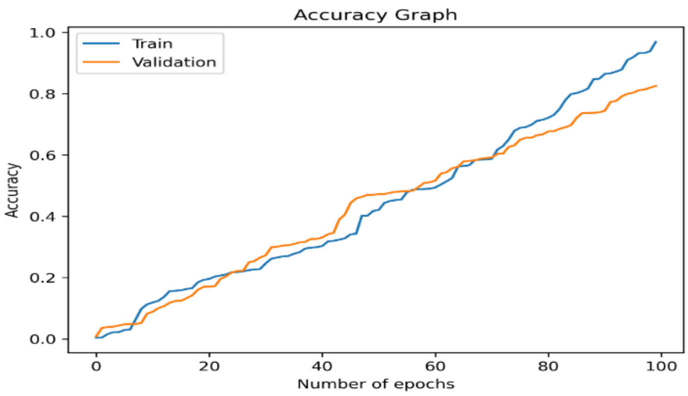

Fig 7: The suggested model's total accuracy, which is 98% accurate during training and 82% accurate during validation

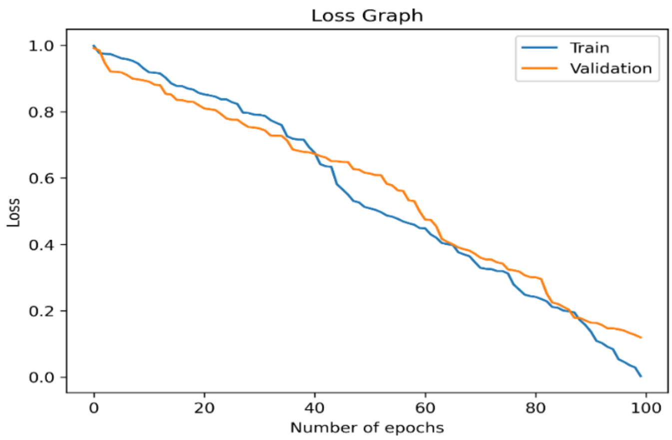

Fig 8: The overall loss percentage of the proposed model with 0% of training loss and 20% of validation loss.

The author had only previously provided the results of the suggested model; henceforth, a comparative analysis would be employed to assess the proposed model's outcomes in comparison to those of the previous study model. In this scenario, the criteria that were applied to generate each comparative analysis graph correspond to those in the suggested model. Starting here, the author explains how the base model and the recommended model select the data from their respective databases based on three parameters. The graph in Figure 9 shows different values according to the three factors. This result clearly shows that there are more questions in the data of the suggested model. Currently, the author would present the findings from figures 10, 12, and 15, which all indicate the same thing. The percentage of the proposed model's outcomes that are superior to the percentage of all the results of the base model is shown by the metrics recall, precision, and F1-score.

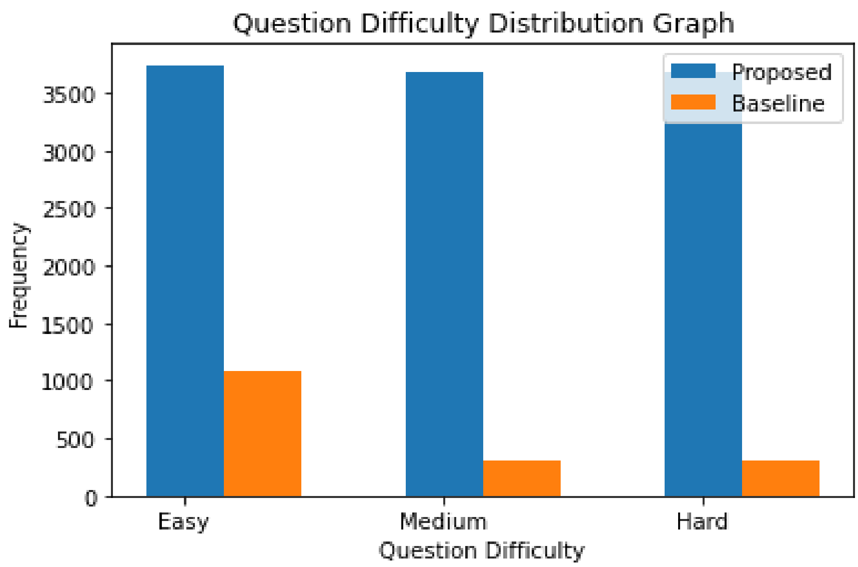

Fig 9: Question Difficulty Distribution Graph

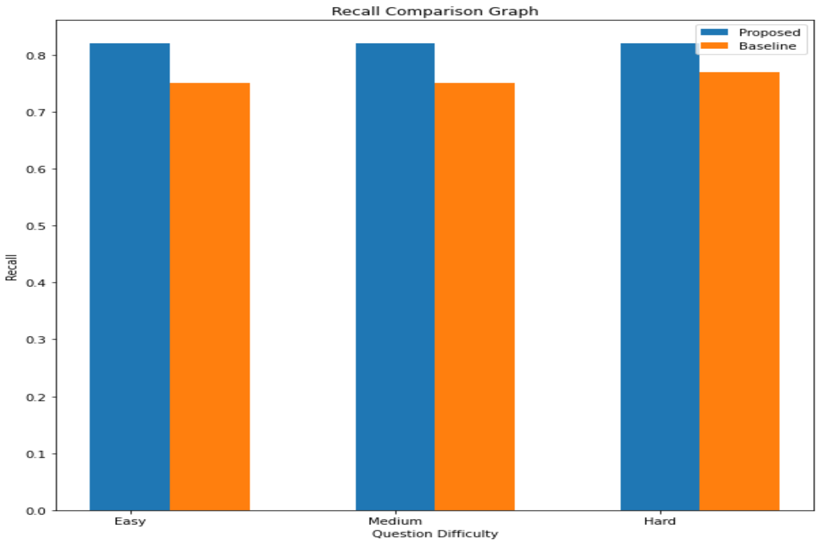

Fig 10: Recall Graph of Comparison.

These three results, which are shown in Figures 10, 12, and 15, according to the three parameters on which the level of the question has been checked. In them, the question level percentage of each parameter is different and the question level percentage of the proposed model is more than that of the base model.

Now, the most significant comparison is shown in the comparative analysis, namely the comparison between the accuracy, loss, and micro average comparison graphs shown in Figures 11, 13, and 14. In this study, the proportion of recommended model outcomes is greater than that of the base model on all parameters. Additionally, the accuracy and loss percentage are superior to the base model. The micro average comparison graph displays the percentage of questions for all parameters, revealing that the suggested model includes varied percentages of easy, medium, and hard questions, and that, based on Figure 14, the proposed model is also superior to the base model.

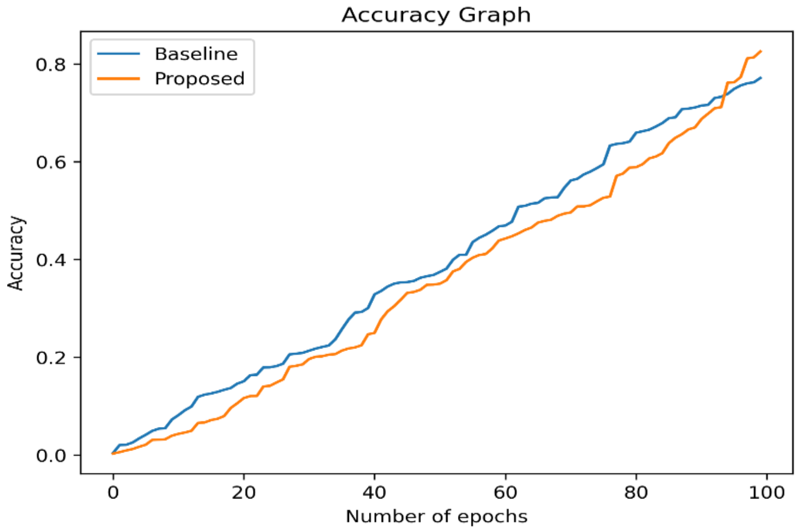

Fig 11: Accuracy Graph of Comparison.

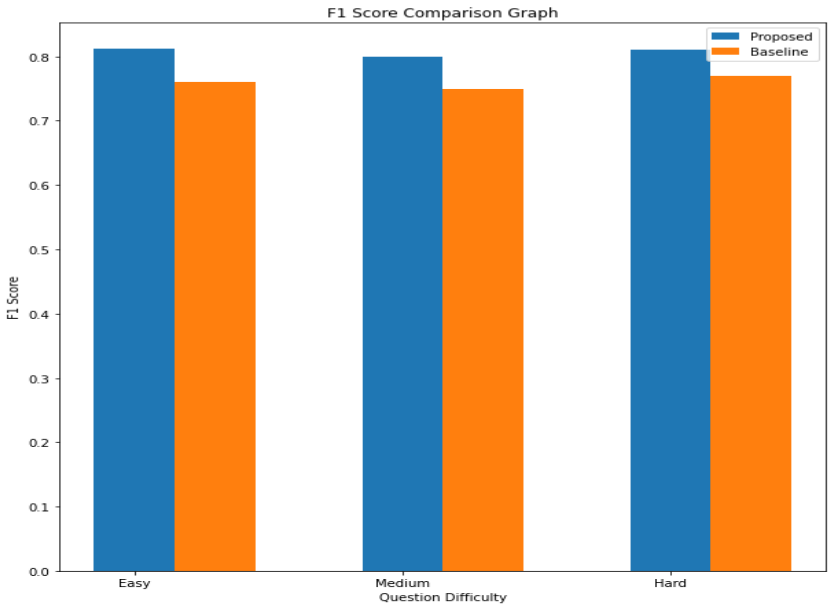

Fig 12: F1-score graph of comparison.

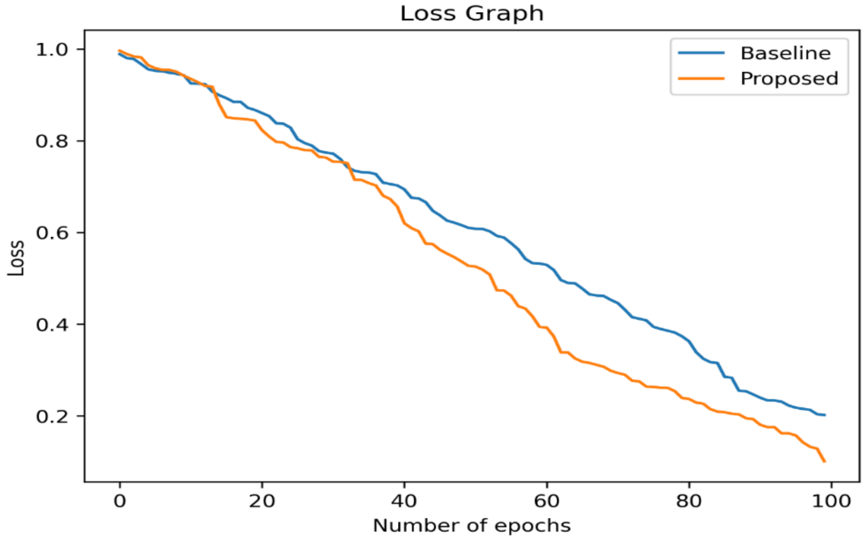

Fig 13: Loss graph of comparison.

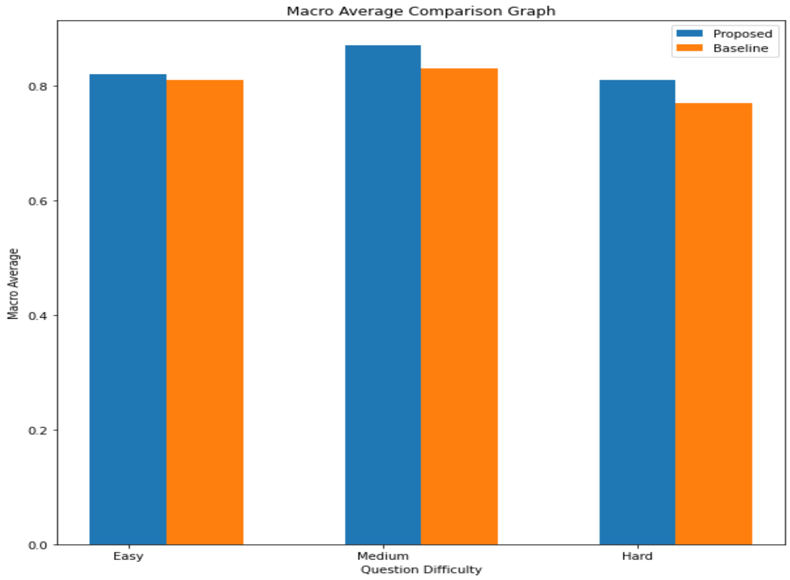

Fig 14: Micro Average Graph of Comparison.

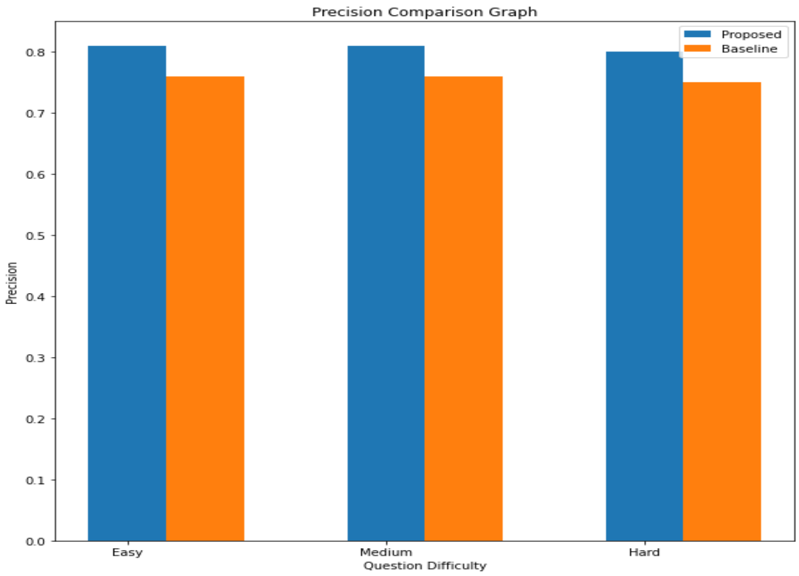

Fig 15: Precision graph of comparison.

Table 3 depict the comparative results as seen below. There are four parameters used such as recall, F1-score, micro average, and precision. In this table 3, question difficulty is estimated in terms of easy, medium, and easy.

From the table 3, our proposed method gives outperformance in every parameter as compared to baseline as seen in table 3.

Table 3: comparative results.

Model	**Parameter**											
	Recall			**F1-score**			**Micro average**			**Precision**		
	Easy	Difficulty Medium Question	Hard	Easy	Difficulty Medium question	Hard	Easy	Difficulty Medium question	Hard	Easy	Difficulty Medium question	Hard
Baseline	0.75 %	0.75%	0.75 %	0.77 %	0.77%	0.77 %	0.78 %	0.80%	0.72 %	0.76 %	0.77%	0.77 %
Proposed	0.80 %	0.80%	0.80 %	0.82 %	0.82%	0.82 %	0.81 %	0.86%	0.80 %	0.80 %	0.80%	0.80 %

6. Conclusion

A BiLSTM model for predicting cognitive processes is proposed by the study. In this experiment, any course questionnaire was considered for analytic purposes. It has been shown that the proposed BiLSTM model accurately predicts cognitive processes such as Understand, Remember, and Apply with an accuracy of 82%. With a training accuracy of 98%, BiLSTM with CNN model feature extraction exceeds previous research in the instance of prediction. Similarly, to the testing phase, the BiLSTM model achieved 82% accuracy. BiLSTM has shown greater performance in the case of a loss compared to preceding investigations. As a consequence, BiLSTM models provide good results.

It helps educators in detecting which cognitive process and component of knowledge students lack. It assists students in dealing with the troublesome skill and recommends that colleges adopt a policy to develop question papers with varying degrees of difficulty like medium,easy, and tough, based on the categories of the questions.

Future evaluations of additional courses from different schools might expand the categorisation relating knowledge dimension and cognitive process. Consideration of the last category of the knowledge dimension, namely the Meta-cognitive domain, might contribute to the development of the research.

References

[1] Dorça, Fabiano A., Luciano V. Lima, Márcia A. Fernandes, and Carlos R. Lopes. "Comparing strategies for modeling students learning styles through reinforcement learning in adaptive and intelligent educational systems: An experimental analysis." *Expert Systems with Applications* 40, no. 6 (2013): 2092-2101.

[2] Abduljabbar, Dhuha Abdulhadi, and Nazlia Omar. "Exam questions classification based on Bloom's taxonomy cognitive level using classifiers combination." *Journal of Theoretical and Applied Information Technology* 78, no. 3 (2015): 447.

[3] Hoque, M. Enamul. "Three domains of learning: Cognitive, affective and psychomotor." *The Journal of EFL Education and Research* 2, no. 2 (2016): 45-52.

[4] Munir, Muhammad Tajammal, Saeid Baroutian, Brent R. Young, and Susan Carter. "Flipped classroom with cooperative learning as a cornerstone." *Education for chemical engineers* 23 (2018): 25-33.

[5] Saido, Gulistan Mohammed, Saedah Siraj, Abu Bakar Bin Nordin, and Omed Saadallah Al_Amedy. "Higher order thinking skills among secondary school students in science learning." *MOJES: Malaysian Online Journal of Educational Sciences* 3, no. 3 (2018): 13-20.

[6] Nidaa, Ghalib Ali, and Salih Hammad Dhiyaa. "CLASSIFYING EXAM QUESTIONS BASED ON BLOOM'S TAXONOMY USING MACHINE LEARNING APPROACH." In *Технологии разработки информационных систем ТРИС-2019*, pp. 260-269. 2019.

[7] Lucas, Kirsten C., Susan M. Dippenaar, and Pieter Hertzog Du Toit. "Analysis of assessment practice and subsequent performance of third year level students in natural sciences." *Africa Education Review* 11, no. 4 (2014): 563-583.

[8] Husain, Farhat N. "Use of Digital Assessments How to Utilize Digital Bloom to Accommodate Online Learning and Assessments?." *Asian Journal of Education and Training* 7, no. 1 (2021): 30-35.

[9] Tanalola, H. S., S. Fattahb, S. R. Sulong, and Mazlina Mamat. "Mining Exam Question based on Bloom's Taxonomy." *School of Engineering and Information Technology, University Malaysia* (2017): 424-427.

[10] Mohd Effendi Ewan Mohd, Matore. "Rasch model assessment for bloom digital taxonomy applications." *Computers, Materials, & Continua* (2021): 1235-1253.

[11] Kamlasi, Imanuel. "Descriptive Analyses on English Test Items based on the Application of Revised Bloom's Taxonomy." *Metathesis: Journal of English Language, Literature, and Teaching* 2, no. 2 (2018): 203-210

[12] Laddha, Manjushree D., Varsha T. Lokare, Arvind W. Kiwelekar, and Laxman D. Netak. "Classifications of the Summative Assessment for Revised Blooms Taxonomy by using Deep Learning." *arXiv preprint arXiv:2104.08819* (2021).

[13] Sami, Joy Christy Antony, and Umamakeswari Arumugam. "A descriptive analysis of students learning skills using bloom's revised taxonomy." *Journal of Computer Science* 16, no. 2 (2020): 183-193.

[14] Kumara, B. T. G. S., A. Brahmana, and Incheon Paik. "Bloom's taxonomy and rules based question analysis approach for measuring the quality of examination papers." *International Journal of Knowledge Engineering* 5, no. 1 (2019): 2-6.

[15] Ullah, Zahid, Adidah Lajis, Mona Jamjoom, Abdulrahman H. Altalhi, Jalal Shah, and Farrukh Saleem. "A rule-based method for cognitive competency assessment in computer programming using Bloom's taxonomy." *IEEE Access* 7 (2019): 64663-64675.

[16] Ali, Nidaa Ghalib, and Dhiyaa Salih Hammad. "CLASSIFYING EXAM QUESTIONS BASED ON BLOOM'S TAXONOMY USING MACHINE LEARNING APPROACH."

[17] Ullah, Zahid, Adidah Lajis, Mona Jamjoom, Abdulrahman Altalhi, and Farrukh Saleem. "Bloom's taxonomy: A beneficial tool for learning and assessing students' competency levels in computer programming using empirical analysis." *Computer Applications in Engineering Education* 28, no. 6 (2020): 1628-1640.

[18] Prasad, G. N. R. "Identification of Bloom's Taxonomy level for the given Question paper using NLP Tokenization technique." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12, no. 13 (2021): 1872-1875.

[19] Makhlouf, Karima, Lobna Amouri, Nada Chaabane, and EL-Haggar Nahla. "Exam Questions Classification Based on Bloom's Taxonomy: Approaches and Techniques." In *2020 2nd International Conference on Computer and Information Sciences (ICCIS)*, pp. 1-6. IEEE, 2020.

[20] Atachiants, Roman, Gavin Doherty, and David Gregg. "Parallel performance problems on shared-memory multicore systems: taxonomy and observation." *IEEE Transactions on Software Engineering* 42, no. 8 (2016): 764-785.

[21] Tanalola, H. S., S. Fattahb, S. R. Sulong, and Mazlina Mamat. "Mining Exam Question based on Bloom's Taxonomy." *School of Engineering and Information Technology, University Malaysia* (2017): 424-427.

[22] Uma, D., S. Thenmozhi, and Rabin Hansda. "Analysis on cognitive thinking of an assessment system using revised Bloom's taxonomy." In *2017 5th IEEE International Conference on MOOCs, Innovation and Technology in Education (MITE)*, pp. 152-159. IEEE, 2017.

[23] Bindra, Simranjeet Kour, Akshay Girdhar, and Inderjeet Singh Bamrah. "Outcome based predictive analysis of automatic question paper using data mining." In *2017 2nd International Conference on Communication and Electronics Systems (ICCES)*, pp. 629-634. IEEE, 2017.

[24] Jayakodi, Kithsiri, Madhushi Bandara, Indika Perera, and Dulani Meedeniya. "WordNet and Cosine Similarity based Classifier of Exam Questions using Bloom's Taxonomy." *International Journal of Emerging Technologies in Learning* 11, no. 4 (2016).

[25] Patil, Soumya K., and M. M. Shreyas. "A comparative study of question bank classification based on revised bloom's taxonomy using svm and k-nn." In *2017 2nd International Conference On Emerging Computation and Information Technologies (ICECIT)*, pp. 1-7. IEEE, 2017.

[26] Fayyad, Usama M., Gregory Piatetsky-Shapiro, and Padhraic Smyth. "Knowledge Discovery and Data Mining: Towards a Unifying Framework." In *KDD*, vol. 96, pp. 82-88. 1996.

[27] Rotondo, Anna, and Fergus Quilligan. "Evolution paths for knowledge discovery and data mining process models." *SN Computer Science* 1, no. 2 (2020): 1-19.

[28] Rogalewicz, Michał, and Robert Sika. "Methodologies of knowledge discovery from data and data mining methods in mechanical engineering." *Management and Production Engineering Review* (2016).

[29] Yates, Darren, and Md Zahidul Islam. "Data Mining on Smartphones: An Introduction and Survey." *ACM Computing Surveys (CSUR)* (2022).

[30] Husain, Farhat N. "Use of Digital Assessments How to Utilize Digital Bloom to Accommodate Online Learning and Assessments?." *Asian Journal of Education and Training* 7, no. 1 (2021): 30-35.

[31] Tanalola, H. S., S. Fattahb, S. R. Sulong, and Mazlina Mamat. "Mining Exam Question based on Bloom's Taxonomy." *School of Engineering and Information Technology, University Malaysia* (2017): 424-427.

[32] Mohd Effendi Ewan Mohd, Matore. "Rasch model assessment for bloom digital taxonomy applications." *Computers, Materials, & Continua* (2021): 1235-1253.

[33] Bhargav, H. S., Gangadhar Akalwadi, and Nitin V. Pujari. "Application of blooms taxonomy in day-to-day examinations." In *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, pp. 825-829. IEEE, 2016.

[34] Sharfuddin, Abdullah Aziz, Md Nafis Tihami, and Md Saiful Islam. "A deep recurrent neural network with bilstm model for sentiment classification." In *2018* International *conference on Bangla speech and language processing (ICBSLP)*, pp. 1-4. IEEE, 2018.

[35] Ms. Pooja Sahu. (2015). Automatic Speech Recognition in Mobile Customer Care Service. International Journal of New Practices in Management and Engineering, 4(01), 07 - 11. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/34

[36] Mondal , D. . (2021). Remote Sensing Based Classification with Feature Fusion Using Machine Learning Techniques. Research Journal of Computer Systems and Engineering, 2(1), 28:32. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/16