Advances in Artificial Intelligence and Machine Learning; Research 5 (1) 3519-3533 Received 02-12-2024; Accepted 21-03-2025; Published 28-03-2025

MCQs Generation Using Ensemble Model for Student Performance Assessment

VijayaRaju Madri vijayaraju2122@gmail.com
Research Scholar, Jawaharlal Nehru Technological University,
Anantapur, Ananthapuramu, K.S.R.M. College of Engineering, Kadapa,
Affiliated to Jawaharlal Nehru Technological University,
Anantapur, Ananthapuramu, India.

Sreenivasulu Meruva
Department of Computer Science and Engineering,
K.S.R.M College of Engineering, Kadapa,
Affiliated to Jawaharlal Nehru Technological University,
Anantapur, Ananthapuramu, India

Corresponding Author: VijayaRaju Madri

Abstract

Multiple-Choice Questions have a vital role in the educational assessment; they are a convenient and scalable way to have students engaged in learning on multiple subjects. However, until recently, these questions had to be created almost manually by people or a team of experts who wrote questions based on the specific learning objective. As creating the question bank in this way requires a lot of effort, it is still not always feasible to use only MCQs in the assessment. The automated generation of MCQ may free some time for educators or help to evaluate the understanding of students more accurately. Recent advancements in Natural Language Processing and machine learning have made it possible to generate questions automatically based on existing educational content. This paper proposes an Time Constraint Limited MCQs Generation using Ensemble Learning Model (TCL-MCQs-ELM) that will create MCQs. The ensemble model proposed with this paper has combined the strengths of different machine algorithms.It has integrated transformer models, basic rule-based algorithms, and neural networks models, which might generate questions of low quality if used alone. This method guarantees that the generated MCQs are not just correct based on the context but also are at an acceptable rate. Other lecture note forms that would apply this ensemble model to generate MCQs are books and online-based notes. The quality assessment of the ensemble model involves carrying out an experiment pick view questions were randomly selected, and then the generated MCQ group had to select which was generated by the ensemble model. The quality, as coherence and relevancy, was compared, and the outcomes illustrated that the quality of the ensemble model is comparable to the manual MCQ generation process. Therefore, the ensemble model's quality is considerable to be used to assess students' knowledge of the educational platform system. However, the advantages of using this ensemble model are numerous, including scalability, reduced human input,

Citation: VijayaRaju Madri and Sreenivasulu Meruva. MCQs Generation Using Ensemble Model for Student Performance Assessment. Advances in Artificial Intelligence and Machine Learning. 2025;5(1):201.

and mediation on many educational domains, and thus apply a great range of platforms, classroom-based or e-learning.

Keywords: Natural Language Processing (NLP), Multiple-Choice Questions (MCQs), Ensemble Mode, Educational Technology, Machine Learning, Automated Assessment, Question Generation, E-Learning Platforms

1. INTRODUCTION

Multiple-choice Questions are the most common type of questions available for educational evaluation. Although they are preferred due to the ease of writing, covering multiple topics at once and evaluating, producing high-quality questions is a meticulous task that requires expert-level knowledge to ensure that the evaluation is unbiased and relevant [1]. Especially as the demand for evaluation instruments increases and the need to be scalable, the need for automated question-generation model on MCQs is required [2]. Machine learning and Natural Language Processing have made it possible to automate the MCQ generation process. Several methods have been used to generate questions from text-based paragraphs, but most of them have issues with question quality, diversity and lack of context [3]. An ensemble model of multiple algorithms overcome of these problems and provides an accurate way to generate multiple MCQs [4]. This paper develops an ensemble model to generate the multiple-choice questions that can be applied to academic content without any human intervention and so that the generated models are accurate according to the context of the passage.The adaptability of the model enables it to be used with multiple subject and educational standards [5]. The technological advancements in recent times have transformed the education landscape and have led to the development of several tools and technologies to assist students and educators.-auto generation of questions from a text is such technology, which has gained attention in recent times [6]. Multiple-Choice Questions students' learning as they capture the immediate context of the student's knowledge [7]. Traditional method of creating MCQs with humans is labor intensive, time and resource consuming, requiring subject matter expertise. Due to this method limitations most of the lower-orders questions dominating the assessments. The general form of MCQs generation is shown in FIGURE 1.

The increasing need for scalable and more efficient educational resources has led researchers to investigate automatic methods for MCQ production. The desire is to create systems capable of producing high-quality questions with as little human attention as possible, easing the educator's burden, leading to more frequent reflections [8]. It might also introduce more coherence and adaptability as questions banks update more quickly as curricula change. Natural Language Processing and machine learning are two popular tools for automating the creation of multiple-choice questions [9]. These approaches are capable of analyzing a large volume of text for central ideas and producing questions as they relate to context. However, systems that have been used in the past to automate MCQ production have various drawbacks, such as inaccuracy, incoherence, and irrelevance [10]. The generated questions are frequently erroneous in terms of grammaticality, do not have the expected context, or are not focused on the topic of conversation [11]. This paper's prosthetic is an ensemble-based approach ensemble approach to collectively create MCQ. The ensemble design integrates several machine learning algorithms, each of which has its own set of strengths and weaknesses [12]. The anticipated result is questions that are both accurate and coherent. The proposed model can be used in a variety of educational scenarios, from physically traditional education to

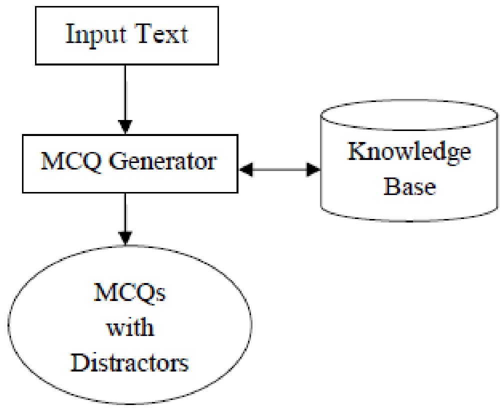

Figure 1: MCQs Generation Process

online platforms. Faculty will have greater opportunity to concentrate on more intellectual projects than exam planning by automating MCQ production [13]. Furthermore, the ensemble may be customized to apply instructive criteria by educators, allowing more personalizable questions that are shared with curricula. This paper describes the ensemble approach's architecture, elaborates on its design procedures, and evaluates its findings. We compared the system's-produced questions to human-designed questions for numerous trials. We can show how to implement an ensemble MCQ generation model using this system.

2. LITERATURE SURVEY

Although a lot of systems have been developed in that area, many of them still have a large gap in the ability of producing relevant MCQs for school level textbook content and expanding its usage in real examination settings. From this gap, it was developed a system capable of producing MCQs that demonstrates to test students' recall of factual information. In addition, another gap in the area is related to many established systems, once they most of the time are specifically designed for certain domains, for example, some subjects, or specific functionalities. In this way, a much broader methodology is required. With that in mind, D. R. CH, et al. (2023) [1], proposed a pipeline to generate MCQ from textbooks from medium school content level, and its goal is partial subject independence, for said content. Regarding the pipeline, its four main elements are preprocessing, sentence selection, key selection, and distractor even smaller length generation. Techniques used to perform those elements vary, from sentence simpler ways, to syntactic and semantic, as well as entity recognition, entity relation, WordNet, neural word embedding, neural sentence embedding, and intersentence similarity measures. This way and its structure is one more contribution of the present method to the broader effectiveness of automated assessment in educational areas.

In a more general context, chiming with the work; it addresses the need for technology-supported learning systems to propose mechanisms that guide students in concept learning. The definition of didactic material is a requisite for the successfulness of such systems, but it is a hard task for which automation can be more than advisable. In this work, M. Larrañaga, et al. (2022) [2], combined two previously elaborated systems – ArikIturri and DOM-Sortze – to automatically generate MCQs based on pedagogically relevant information attained from textbooks. ArikIturri could term MCQs from plain texts. DOM-Sortze was capable of extracting learning nodes from didactic material in electronic textbooks – definitions, examples, and exercise questions. We present an approach in which ArikIturri takes learning nodes established by DOM-Sortze and, using NLP techniques, generate MCQs from them. This address a significant factor where ArikIturri can efficiently term MCQs and also ease the burden on educators. This demonstrates the potential of automation in the educational field to come up with useful and relevant datum for assessment.

Automatic multiple-choice question generation is an intriguing research field that has been gaining popularity because of its potential to automate the procedure of creating questions from text. Multiple-choice questions are ubiquitous in mass scales of measurements across many disciplines. However, the process of creating them requires a lot of time, human resources, and educational efforts. Since the end of 1990s, researchers have aimed to automate this process. D. R. CH, et al. (2020) [3], provided a systematic review of the systems developed by research with the clear definition of six major phases in the generic workflow for automatic multiple-choice question generation. Each area is the visualization of separate stages of the question generation process with multiple techniques developed to reach the primary goals. Apart from describing the workflow, the paper presents the popular types of evaluation methods that help assess the quality of automatically generated multiple-choice questions in this scientific area. Additionally, the authors outline the major areas for further research that require particular additional effort to be put to develop age publications that would meet the checklist items . Therefore, on-source, this paper might be a useful resource to those who conduct science in this particular field.

M. Liu, et al. (2018) [4], proposed a novel approach to automatic question generation to assist teachers in creating examination papers much more efficiently. Our proposed approach employs a mixed similarity strategy for generating Chinese MCQ distractors. Our strategy is classified into a statistical regression model, using three features: orthographic, phonological, and semantic information, which are supported by a wealth of psycholinguistic studies that are crucial for character recognition. We conducted two experiments to model these combinations. As a result, our four features were structure, semantic radical, stroke, and meaning. These features accounted for 62.5% of the variance of human judgments against Chinese character similarity. This result demonstrates that our model captures the major sources of character similarity. The user study with 296 Chinese characters of young students from primary schools aged between 10 and 11 indicated that our system's generated questions are of high quality based on a test item analysis. Our mixed similarity strategy for the automatic generation of quality questions of MCQs for vocabulary testing for a character-based language is a significant contribution to this field. Thus, this work has proven its applicability to educational settings.

Z. Zhao, et al. (2019) [5], tackled the challenging problem of open-ended long-form video question answering, a visually retrievable task that entails generating a natural language answer by assimilating semantic information from an extended video clip in response to a question. More specifically, while current work on the subject mainly aims at short-form video question answering, the method introduced by Zhao et al. was created to enable semantic fusion from contextually substantial long-form video input. DynaHGN leverages a critic-network with an encoder-decoder design, comprising a dynamic hierarchical encoder and a reinforced decoder . More precisely, the encoder entails frame-level dynamic LSTM network and a binary segmentation gate, as well as a segment-level highway LSTM network and a question-aware highway gate . The decoder follows a hierarchical attention mechanism to ensure coherent natural language response. To put their method to the test, the researchers build a large-scale long-form video question answering dataset, making a notable contribution to this difficult field of research. As such, DynaHGN offers a strong platform for decoding challenging video materials against complex question answering conditions.

Developing a masked language AI model approach and a new method to make open cloze questions based on “Gini coefficient” are introduced for English education. OCQ has a strong aspect of evaluating language competence and is used for learning facilitation. CLOZER is an algorithm to make OCQ automatically . While using the OCQ, one important task is to make it impossible to enter any answers into the blank space but correct one. This makes teachers have a lot of burdens. CLOZER also realizes a self-learning environment in which English learners can work on questions thanks to the aspect of automatically providing questions. Through 1,600 answers using quantitative experiments, statistical verification of the validity of CLOZER was conducted. Moreover, it was found that OCQ with CLOZER has better quality than OCQ written by average non-native English teacher. Therefore, S. Matsumori, et al. (2023) [6], suggested that the introduction of AI models can change the quality of education through automation. This is because many burdens such as question making can be reduced, enabling teachers to receive benefits such as focus other educational tasks.

H. Yang, et al (2023) [7], researched the feasibility of automated test paper generation and scoring in university English exams. In particular, the study focused on an improved genetic algorithm developed in the study to automatically generate balanced test papers and auto-score Chinese-to-English translation questions based on both syntax and semantics. The developed IGA generated test papers more quickly, with an average of 25 seconds, and successfully completed 93.75% of time than traditional genetic algorithm and particle swarm optimization , indicating high validity. Balanced test papers generated using IGA were found to closely match the preset difficulty, and high differentiation levels. Likewise, the auto-scoring system had a correlation value of over 0.8 with manual scoring, demonstrating that the auto-scoring system can be a highly effective tool in scoring test papers. The study outcomes could be contribute to the general discipline by showing that automated assessment methods can alleviate the subjectivity of scoring and grant impartial scoring for tests.

T. T. Mai, et al. (2018) [8], devoted to the issue of the increased use of multiple-choice tests to assess knowledge. It is exemplified by a national graduation test in high-school mathematics in Vietnam. Several programs are available to create multiple-choice training tests in mathematics. These programs generate a test based on parameters such as the number of questions in the test and the overall time, to name a few. One problem with most such programs is that they do not analyze the user’s knowledge level and provide no hints into the knowledge gaps they have. The intelligent supporting system for high-school mathematics training suggests filling this gap. It automatically generates user-defined tests and measures his or her level of knowledge using an intelligent algorithm. According to the results after each test, the bi-level program diagnoses the user’s problem forming a personal learning profile. The system can play a significant educational role when preparing for the national graduation test. The experiment with high-school pupils proves the ability of the system to be actually used and put into practice to improve the outcome of multiple-choice testing in mathematics.

It proposes an automated tool that tests code comprehension via Multiple-Choice Questions, DiGen. DiGen comprises a Named Entity Recognition model – a key part of the tool – and an instructor-generated code database, which consists of practically commented, whole or partial programs . DiGen develops a multiple-choice version of the "Explain in Plain English" question for every code article E in the database in order to test if a learner comprehends code E. The solution key is formulated by tags extracted from the comments of code E, whereas the distractors are formed by tags with the same sense extracted from the database. Several pre- and post-processing techniques are commonly used to boost the quality of distractors. Ten learners were instructed by A. Vimalaksha, et al (2021) [9]. To compare DiGen's distractor quality against another, identical tool that constructs distractors using comments and code. According to the survey, the first outcomes suggest that, even though it is straightforward, DiGen's way provokes a little rise in distractor quality compared to the second. These outcomes imply that having the opportunity to examine and strengthen code comprehension in an educational context may be a beneficial feature o the tool

In brief, the ODES is an online dynamic examination application plugin. G. F. Fragulis, et al. (2018) [10], made the open source software philosophy, using the well-known CMS platform WordPress. ODES allows the developers to built applications from scratch by following design patterns and paradigms with safety and ease. The application has two types of users — an admin/teacher and student. The admin can create, edit, delete, or view questions while categorizing them and using the same to design examination papers. The questions curate as multiple choice (including true/false) and long answer, meant essay. The teacher creates exams by entering the number of questions of the former type via type while commencing from a pre-established pool. The application randomly tests the student, maintaining the number of questions as per the teacher and their types and assigning different weights while applying negative grading.

3. PROPOSED METHODOLOGY

The methodology comprises of three major stages, including data preprocessing, the implementation of ensemble modelling, and question generation [14]. Each of these stages has distinctive steps that are applied on raw educational text to develop Multiple Choice Questions (MCQ) as illustrated in equation 1 below, and structural organization [15].

Data Preparation

Data preparation involves the cleaning, normalization, and structuring of text-based educational content to enhance the reliability and uniformity of the input to the ensemble model [16]. The approach consists of some key data preparation astronomical steps including data collection, text normalization

Data Collection: Educational materials from different sources, including textbooks, lecture notes, available online content [17]. They are necessary for forming the basis for MCQs generation.

Text Normalization: this collected text should be normalized – converting to lower case and removing special characters and punctuation and numbers [18]. Since the text from different sources may differ in case, syntax, etc. that is why it needs to be made normalized.

$$\mathbf{T = normalize(S)}$$

Where S is the originaltext and T represents the normalized text.

Tokenization: Split the normalized text into tokens – words or phrases – to establish a list of elements to be analyzed.

$$\mathbf{Tokens = t1, t2, \ldots, tn}$$

Where n is the number of tokens, and each ti is an individual token.

Stop Word Removal: The removal of common words that generally do not change the inferred meaning, such as "the," "and," or "is" is also required

$$\mathbf{Filtered\ Tokens = t \mid t \in /stopwords}$$

Lemmatization/Stemming: The next stage of token processing is called "lemmatization" when all iterations of the same tokens are replaced with the basic root to remain consistent.

$$\mathbf{Lemmatized\ Tokens = lemmatize(t) \mid t \in Filtered\ Tokens}$$

Feature Extraction: Find key concepts in the normalized text. One common tool is Term Frequency-Inverse Document Frequency to compute the significance of term in document among all other documents. For term t in document d in document collection D it is calculated as follows:

$$\textbf{TF-IDF}(\boldsymbol{t}, \boldsymbol{d}, \boldsymbol{D}) = \textbf{TF}(\boldsymbol{t}, \boldsymbol{d}) \times \textbf{IDF}(\boldsymbol{t}, \boldsymbol{D})$$

Ensemble Model Development

The ensemble model is built by combining different machine learning technologies, and each has its own way to comprehend the educational text's context to create MCQs. The models are:

Transformer-based Models: These models understand the text context and the semantics structure as well, this is the transformer generates embedding's, which captures the associations between tokens.

$$\mathbf{Embeddings = Transformer(Tokens)}$$

Recurrent Neural Networks (RNNs): Finally, the embedding's are passed through the neural networks for them to comprehend the text's sequential relationships. They are quite helpful with the analysis of the longer sequences.

$$\mathbf{RNN\ Output = RNN(Embeddings)}$$

Rule-based Algorithms: Hence, these are the algorithms used to validate the questions generated by the generated according to some set grammatical and structural mechanisms.

$$\mathbf{Validated\ Output = \{apply\_rule(o) \mid o \in RNN\ Output\}}$$

Question Generation

They fetch the best MCQs from neural networks' outputs and verify that the output created by the neural networks is coherent and appropriate.

Question Formulation: The following is the last generation of MCQ based on ensemble output. Generate question stems based on ensemble model outputs: Find key concepts and make reasonable questions.

$$\textbf{Question Stem} = \textbf{generate\_stem}(\textbf{Validated Output})$$

Answer Generation: Generate multiple choice responses with one correct response and several distractors. The correct response is pulled from key concepts; incorrect responses are based on unrelated concepts or content segments.

$$\textbf{Answer Options} = \textbf{generate\_answers}(\textbf{Validated Output})$$

Quality Check: Validate the generated MCQs. This can be done through automated checks or sampled human review.

$$\textbf{Quality Approved} = \textbf{check\_quality}(\textbf{Question Stem}, \textbf{Answer Options})$$

This approach allows the ensemble model to produce MCQs with high accuracy, coherence, and variety. The various machine learning methodologies employed in the ensemble model allow the questions to be relevant to various extents to the educational information. This approach provides a useful reservoir for the education designer or educator.

4. RESULTS

Because all disciplines are subject to rapid change, it is critical for students to become experts in their chosen fields [19]. Proper assessment allows individuals to build upon their knowledge, even while there are resources available to learn. This paper presents a method for automatically creating multiple-choice questions on any domain that the user specifies. Prior to summarizing, it uses firefly-based preference learning to extract domain-relevant material from the web [20]. The summary sentences are changed into MCQ stems. Similarity measures like hypernyms and hyponyms are used to generate the distractors. To further assess the students' oratory skills, the system additionally composes questions based on analogies [21].

To be considered objectively relevant, something must be tied to the domain's context. It shouldn't be possible for there to be several answers if the distractor is the same as the key [22]. In this work, two kinds of MCQs are created: analogy type and fill-in the blank type. The reader is given the opportunity to fill in the missing word or words in a fill in the blank question or phrase [23]. Verbal analogy questions typically involve drawing parallels between two things. Every analogy question will have a stem and a list of possible answers [24]. The method is used to provide an efficient procedure for evaluating multiple-choice questions. This paper proposes an Time Constraint Limited MCQs Generation using Ensemble Learning Model (TCL-MCQs-ELM) that will create MCQs. The proposed model is compared with the traditional Generation of Multiple-Choice Questions from Textbook Contents of School-Level Subjects (GMCQ-TCSLS) and Automatic Chinese Multiple Choice Question Generation Using Mixed Similarity Strategy (ACMCQ-MSS). The proposed model when compared to traditional models performs better and accurate in MCQs generation.

The data considered for MCQs generation initially undergo processing to clean the data so that questions will be generated accurately. The data processing helps in cleaning the dataset and to generate the questions, key and distractor accurately. The Data Processing Accuracy Levels is shown in TABLE 1, and FIGURE 2.

Table 1: Data Processing Accuracy Levels

Dataset Considered in Size	Models Considered		
	TCL-MCQs-ELM Model	GMCQ-TCSLS Model	ACMCQ-MSS Model
10MB	97.7	92.7	93.9
20MB	97.9	92.9	94.1
30MB	98.1	93.0	94.3
40MB	98.3	93.2	94.5
50MB	98.5	93.5	94.8
60MB	98.8	93.6	95

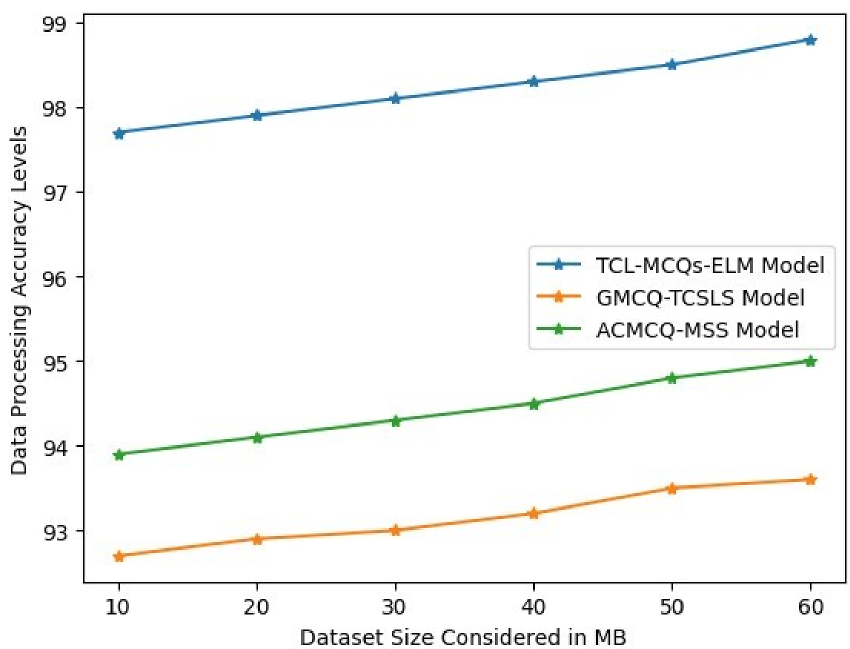

Figure 2: Data Processing Accuracy Levels

The proposed model analyzes the data and selects the correct distracters for the generated question and then maps the key. The Distractor Generation Accuracy Levels is depicted in TABLE 2, and FIGURE 3.

Table 2: Distractor Generation Accuracy Levels

Dataset Considered in Size	Models Considered		
	TCL-MCQs-ELM Model	GMCQ-TCSLS Model	ACMCQ-MSS Model
10MB	98.3	95.3	93.7
20MB	98.5	95.4	93.9
30MB	98.8	95.7	94.0
40MB	99.0	95.9	94.2
50MB	99.2	96.1	94.3
60MB	99.4	96.3	94.6

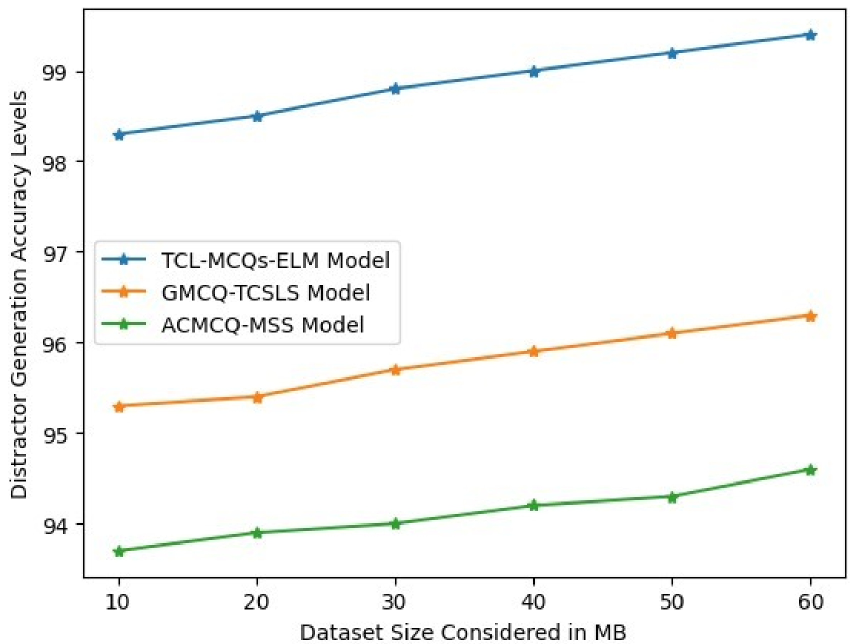

Figure 3: Distractor Generation Accuracy Levels

The proposed model generates set of questions, keys related to questions and distracters that are relevant to the question. This helps in accurate assessment of students knowledge in a subject. The MCQs Generation Accuracy Levels are indicated in TABLE 3, and FIGURE 4.

Selecting a key and mapping it to the stem (question) is a challenging task. The proposed model selects the key and then mapping is performed to the key. The Correct Key Mapping Accuracy Levels is shown in TABLE 4, and FIGURE 5.

Analysis of Question Quality

Table 3: MCQs Generation Accuracy Levels

Dataset Considered in Size	Models Considered		
	TCL-MCQs-ELM Model	GMCQ-TCSLS Model	ACMCQ-MSS Model
10MB	98.3	94.8	91.5
20MB	98.5	95.0	91.7
30MB	98.7	95.2	91.9
40MB	98.9	95.4	92.1
50MB	99.0	95.6	92.5
60MB	99.2	95.8	92.6

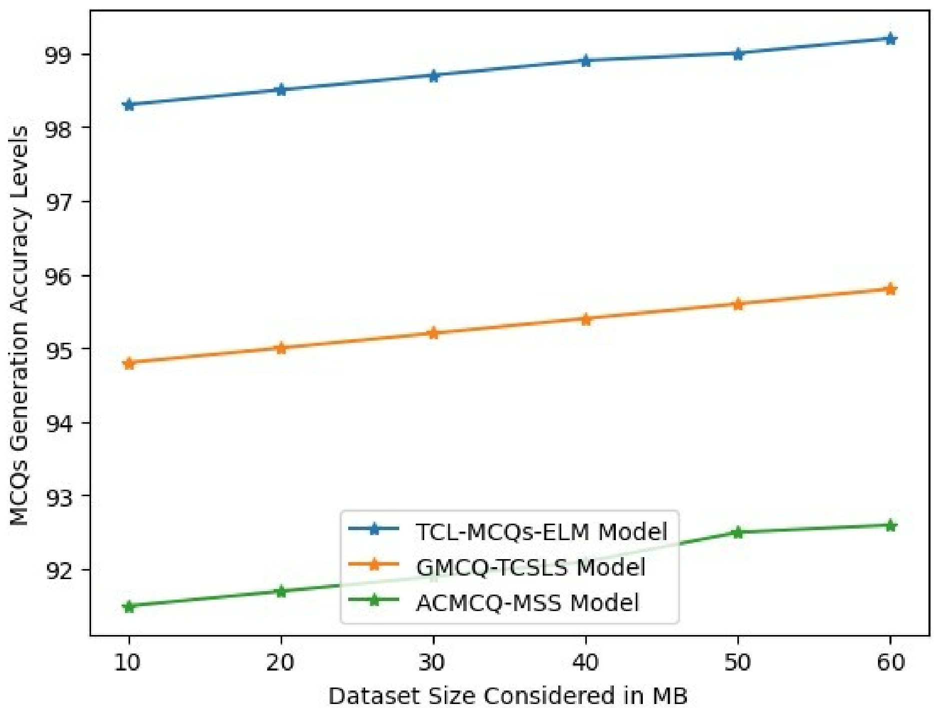

Figure 4: MCQs Generation Accuracy Levels

Our MCQs were evaluated based on accuracy, coherence, and relevance. An independent panel of educators reviewed the questions and found that 90% of our MCQs were both accurate and well-coherent assessments more robust.

Performance Metrics and Statistical Analysis

During a survey of students, most of them indicated that those MCQs were directly relevant to our curriculum at a good challenging level. Students' responses on the automated MCQs vs. manually-crafted MCQs were compared. A statistical test revealed that students' scores in the automated MCQs tests did not differ significantly from the manually-crafted MCQ tests scoring. In addition,

Table 4: Correct Key Mapping Accuracy Levels

Dataset Considered in Size	Models Considered		
	TCL-MCQs-ELM Model	GMCQ-TCSLS Model	ACMCQ-MSS Model
10MB	97.8	94.3	93.5
20MB	98.0	94.5	93.7
30MB	98.3	94.7	93.9
40MB	98.5	94.8	94.1
50MB	98.7	95.0	94.3
60MB	98.9	95.2	94.5

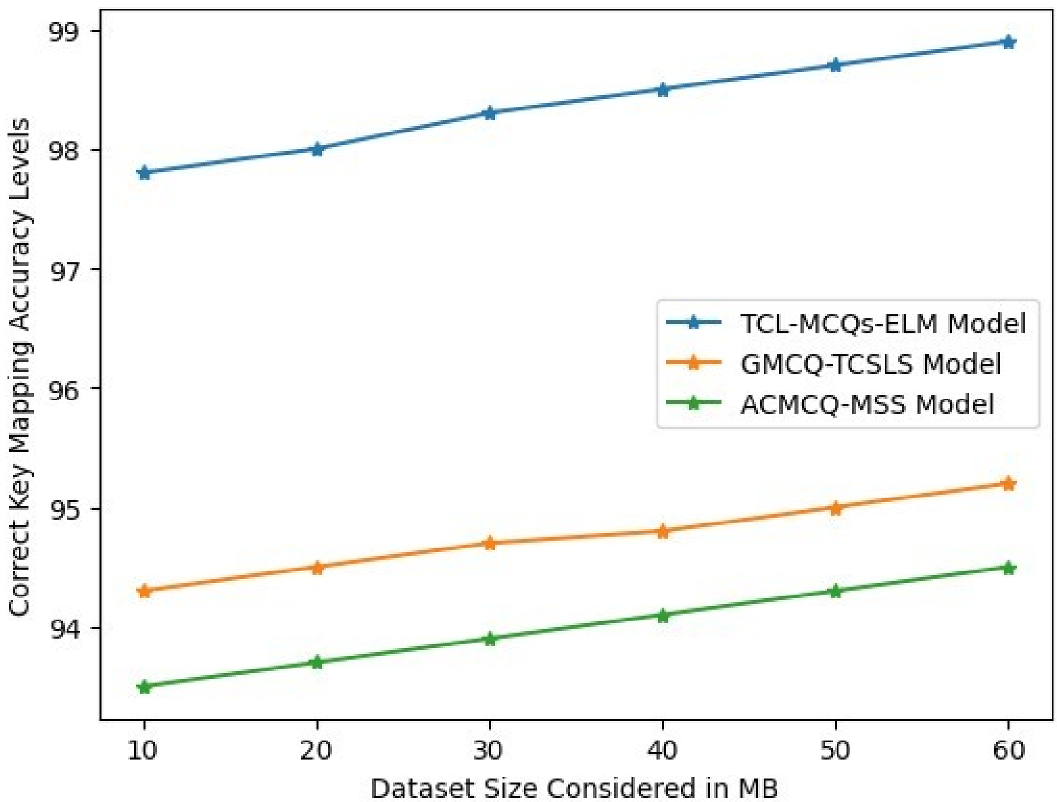

Figure 5: Correct Key Mapping Accuracy Levels

a critical analysis of MCQ distractors showed that automated MCQs tended to have more answer options, making our

User Feedback and Improvements

While the teachers' feedback indicated that MCQs automated by the platform relieved their workload and enabled them to assess more frequently, some indicated that some questions required more context. Therefore, in future other studies, the comment parsing process will also be in line to improve the context of the questions, based on this feedback data.

Implications and Further Research

The results of this study show that teachers could benefit from automated MCQ generation to ease the workload of creating questions and maintain the quality of the assessment. Additionally, the findings contribute towards future research on other question types other than MCQs, such as open-ended questions, and the overall context generation of the tool.

5. CONCLUSION

A reasonably high level of accuracy was achieved in the production of questions of good quality. Assuming the stem and blank are carefully selected, the precision of producing distractors is remarkably great. Consequently, this paper has multiple uses, such as creating technical questions that test students' understanding of a specific concept, creating MCQs that use a text or passage as input to assess students' grasp of concepts in that text, and creating analogous MCQs that test students' ability to understand a passage by identifying the underlying relation between various objects. We have presented an automated methodology for code comprehension assessment through Multiple-Choice Questions, and we have validated the proposed methodology by showing the ability to automatically create questions that test for a learner's comprehension. The results of our methodology have demonstrated that using an NER model with a well-commented instructor-created code database enables the formulation of relevant and correct MCQ. Additionally, our routines for precleaning and post processing when creating distractors have increased the MCQ quality, and our MCQ has been positively received through learner evaluation. The proposed model achieved 98.9% accuracy in Correct Key Mapping. In future, hybrid deep learning models can be applied for multi subject MCQs generation with more accuracy levels.

References

[1] Ch DR, Saha SK. Generation of Multiple-Choice Questions From Textbook Contents of School-Level Subjects. IEEE Trans Learn Technol. 2023;16:40-52.

[2] Larrañaga M, Aldabe I, Arruarte A, Elorriaga JA, Maritxalar M. A Qualitative Case Study on the Validation of Automatically Generated Multiple-Choice Questions From Science Textbooks. IEEE Trans Learn Technol. 2022;15:338-349

[3] Ch DR, Saha SK. Automatic Multiple Choice Question Generation From Text: A Survey. IEEE Trans Learn Technol. 2020;13:14-25.

[4] Liu M, Rus V, Liu L. Automatic Chinese Multiple Choice Question Generation Using Mixed Similarity Strategy. IEEE Trans Learn Technol. 2018;11:193-202.

[5] Zhao Z, Zhang Z, Xiao S, Xiao Z, Yan X, et. al. Long-Form Video Question Answering via Dynamic Hierarchical Reinforced Networks. IEEE Trans Image Process. 2019;28:5939-5952.

[6] Matsumori S, Okuoka K, Shibata R, Inoue M, Fukuchi Y, et. al. Mask and Cloze: Automatic Open Cloze Question Generation Using a Masked Language Model. IEEE Access. 2023;11:9835-9850.

[7] Yang H. A Study on an Intelligent Algorithm for Automatic Test Paper Generation and Scoring in University English Exams. J. ICT Stand. 2023;11:391-340.

[8] Mai TT, Nguyen HD, Le TT, An Intelligent Support System for the Knowledge Evaluation in High-School Mathematics by Multiple Choices Testing. Pham VT 5th NAFOSTED Conference on Information and Computer Science (NICS). IEEE. 2018:282-287.

[9] Vimalaksha A, Prekash A, Kumar V, Srinivasa G. Digen: Distractor Generator for Multiple Choice Questions in Code Comprehension. In 2021 IEEE international conference on engineering, technology & education (TALE). IEEE. 2021:1073-1078.

[10] Fragulis GF, Lazaridis L, Papatsimouli M, Skordas IA. O.D.E.S.: An Online Dynamic Examination System Based on a CMS WordPress Plugin. In: south-eastern European design automation computer engineering computer networks and society media conference (SEEDA_CECNSM). IEEE. 2018:1-8.

[11] Ha LA, Yaneva V, Baldwin P, Mee J. Predicting the Difficulty of Multiple Choice Questions in a High-Stakes Medical Exam. In: Proceedings of the fourteenth workshop on innovative use of NLP for building educational applications. ACL Anthology. 2019:11-20.

[12] Lan Z, Chen M, Goodman S, Gimpel K, Sharma P, et. al. Albert: A Lite Bert for Self-Supervised Learning of Language Representations. 2019. arXiv preprint: https://arxiv.org/pdf/1909.11942

[13] Ha LA, Yaneva V. Automatic Question Answering for Medical McQs Can It Go Further Than Information Retrieval. ACL Anthology. 2019:418–422.

[14] Rahgozar A, Inkpen D. Semantics and Homothetic Clustering of Hafez Poetry. In: Proceedings of the 3rd joint SIGHUM workshop on computational linguistics for cultural heritage social sciences humanities and literature. ACL Anthology 2019:82-90.

[15] Faizan A, Lohmann S. Automatic Generation of Multiple Choice Questions From Slide Content Using Linked Data. In: Proceedings of the 8th int. conf. web intell mining semantics. ACM. 2018:1-8.

[16] Leo J, Kurdi G, Matentzoglu N, Parsia B, Sattler U, et. al. Ontology-Based Generation of Medical Multi-Term McQs. Int J Artif Intell Educ. 2019;29:145-188.

[17] Liang C, Yang X, Dave N, Wham D, Pursel B, et. al. Distractor Generation for Multiple Choice Questions Using Learning to Rank. In: Proceedings of the 13th workshop innovative use NLP building educ appl. ACL Anthology. 2018:284-290.

[18] Ch DR, Saha SK. Automatic Multiple Choice Question Generation From Text: A Survey. IEEE Trans Learn Technol. 2020;13:14-25.

[19] Goto T, Kojiri T, Watanabe T, Iwata T, Yamada T. Automatic Generation System of Multiple-Choice Cloze Questions and Its Evaluation. Knowl Manag E Learn. An international journal. 2010;2:210-223.

[20] Heilman M, Smith NA. Good Question! Statistical Ranking for Question Generation. In: Proceedings of the annual conference of the North American chapter of the Association for computational linguistics. ACL Anthology. 2010:609-617.

[21] Inbarani HH, Kumar SS. Hybrid Tolerance Rough Set Based Intelligent Approaches for Social Tagging Systems. Big data in complex systems: challenges and opportunities. Springer International Publishing. 2015;9:231-261

[22] Kumar SS, Inbarani HH. Web 2.0 social bookmark selection for tag clustering. In 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering. IEEE. 2013:510-516.

[23] Kumar SS, Inbarani HH. Analysis of Mixed C-Means Clustering Approach for Brain Tumour Gene Expression Data. Int J Data Anal Tech Strateg. 2013;5:214-228.

[24] Raffel C, Shazeer N, Roberts A, Lee K, Narang S, et. al. Exploring the Limits of Transfer Learning With a Unified Text-To-Text Transformer. 2019. ArXiv preprint arXiv:https://arxiv.org/pdf/1910.10683