Digital Handwritten Answer Sheet Evaluation System

Mr.K.S. Manojee
Assistant Professor, Department of IT
Kongunadu College of Engineering and Technology
Trichy, Tamil Nadu, India.
manojee.suri@gmail.com

K. Kannika
Student, Department of IT
Kongunadu College of Engineering and Technology
Trichy, Tamil Nadu, India.
kannikakani674@gmail.com

A. Rasiniya
Student, Department of IT
Kongunadu College of Engineering and Technology
Trichy, Tamil Nadu, India.
rasiniyaa@gmail.com

N P. Rima
Student, Department of IT
Kongunadu College of Engineering and Technology
Trichy, Tamil Nadu, India.
rimaselvarani@gmail.com

Abstract – Computerization of summary takes over the world. It requires a lot of time and energy from school teachers and university faculties. To counter this challenge, our project aims to optimize the assessment process by scanning handwritten responses by students in digital textbooks to meet teachers with predefined model responses. It's the purpose. This was achieved using state-ART technologies such as optical character detection (OCR), natural language processing (NLP), and machine learning algorithms. Advanced Bert bidirectional encoder displays from Transformer Models and Cosinus -similarity Algorithms are used for both accurate and effective assessment of student-provided responses. In contrast to response length, this project maximizes brand allocations in relation to the most important words to save time and effort from teachers and to promote a fair assessment process. Furthermore, this also encourages students to better understand the concept and provide accurate and accurate answers that contribute to the production of fair results and the same results.

Keywords— Optical Character Recognition (OCR), Natural Language Processing (NLP), Machine Learning (ML), BERT (Bidirectional Encoder Representations from Transformers), Cosine Similarity, Text Classification, Tokenization, Artificial Nural Network (ANN), Sentence Embeddings.

I. INTRODUCTION

Automation has come into everything in life and is widely applied in the world today. The easiest way to measure a student's learning ability is to look at their test answers. If the system automates the evaluation part, it will be the most accurate measurement tool. Teachers have to make an effort to read every handwriting and check every character in the current marking system. Teachers have to put in a lot of effort while checking the scripts. The time it takes to score and correct a written answer depends on the number of answer sheets. The average teacher spends 20 to 25 minutes marking answers. After grading 50 answers, the time it takes to grade will be an issue. Currently, students are given grades after evaluation. Grading requires you to enter all grades for each student in your class, and the script must be kept safely. This process also takes quite a while. Delays in manual assessment will make students anxious as it is a lengthy process. Grading masses of handwritten responses is a tedious process that is guaranteed to be error. In an education system, integration of assessment tools helps teachers in various ways. Automation of responses will be more effective on a large scale. Considering the increasing and prioritizing requirements of educational institutions, educational institutions need tools that will help teachers save time and preserve class time. Most educational institutions around the world use pen and paper methods in most exams. This put pressure on teachers to reduce student academic time. The correction is simple and not boring, so evaluation equipment should be used for a single purpose. Building relevant and powerful assessment tools is also a challenge. This is because you need to be able to identify and understand handwriting styles. All students' font styles are different from those of other students. If the manuscript is clear and clean, it is easy for a trained system. This helps students work on their ability to improve their writing quality and feedback from teachers. However, there are other similar tools that deal with grammar, keyword similarity comparison. The developed assessment tool works mainly on handwritten answers. Evaluation of response scripts is an important aspect of student evaluation. To evaluate students, teachers apply different methods: B. Answer short questions, answer descriptive questions, answer multiple choice questions. The application of assessment means in educational systems is from multiple choice questions and short answers. Compared to answering descriptive questions, it is easier and less time consuming to evaluate. $f(x) = \max(0,x)$ where x is the input variable of this function. In the proposed algorithm, the length of the solution is also used as a parameter. The length of the written answer is assumed as the teacher. Teacher entries are considered, but the length of written answers is determined.

A. MOTIVATION

By several key factors, this project was motivated. First, OCR and NLP can save time and effort for grading from manual evaluation. Instead, with this, more time is saved and given to the educators to focus on students' needs with which high personalized learning experiences and academic growth are achieved. NLP-based assessment also ensures the reduction of bias and inconsistencies in subjective grading, thus evaluating students fairly and objectively. Integration of technology with the assessment process also aligns with the ever-increasing trend of digitalization in education as it prepares the students with crucial skills for successful participation in a modern workforce.

B. CONTRIBUTION

The most significant contribution of this study is that it delves into new methods of solving the long-standing problems of marking handwritten response sheets. This project aims to revolutionize conventional assessment procedures and usher in a new era of efficiency and objectivity in education by utilizing natural language processing potential. We aim to clarify the feasibility and efficacy of incorporating NLP techniques into the examination process through thorough experimentation and evaluation. The objective of this research is to offer specific suggestions and best practices for the implementation of NLP empowered assessment tools in classroom environments. In the end, though, we hope to have a positive influence on education and enable teachers and students to thrive in an increasingly technology-dependent society.

C. NEEDS

The requirement for the project stems from a series of immense difficulty and inefficiencies associated with the conventional manual grading process. These challenges impact both educators and students, emphasizing the pressing need for an automated solution. The main drivers of the project's formulation are:

- Labor-Intensive Hand Grading: Hand grading is labor-intensive, particularly when handling a large number of students or complicated tasks. Teachers devote considerable amounts of time personally grading every student's answer, which could be spent more effectively on more influential teaching and educational work.

- Inconsistent Grading Standards: Grading by hand is prone to inconsistency. Various educators could evaluate and interpret student responses differently, resulting in grading differences. These differences can negatively impact the quality and equity of education, since students are not always tested under standardized criteria.

- Subjectivity and Bias: Grader subjectivity and grading biases may affect the objectivity of grades. Teachers can unconsciously prefer specific response formats or writing styles, which creates an unlevel playing field for learners. avoiding such prejudices is important for an unbiased assessment of the work of learners.

- Lack of Timely Feedback: Traditional grading often results in delayed feedback to students. This delay can impede students' capacity to grasp and correct their weaknesses in time. Prompt feedback is are necessary to create a proactive learning atmosphere.

- Scalability: In large student-enrollment institutions, the scalability of manual grading is a growing concern. With the increase in the number of students and assignments, educators' workload increases proportionally. An automated system can manage this scalability more efficiently.

- Adaptability and Personalization: Institutions of learning have different standards and grading measures. A dynamic system able to cater to varying grading parameters and the institution's demands is essential. It fits one size doesn't not fit all, and a solution that can be customized is highly valuable.

The "Digital Handwritten Answer Sheet Evaluation System" responds to these urgent requirements by introduces a new, technology-based method of grading. Automating the process of grading, it streamlines the assessment of student responses, enhances the consistency and objectivity of grading, and gives students immediate feedback. This not only lessens the burden on instructors but also enhances overall quality of education through the establishment of an equitable and effective assessment system.

II. SYSTEM ARCHITECTURE

The system architecture diagram is a structured framework that outlines the flow of data, processes, and components within the system. It integrates various input sources, including handwritten files, Excel sheets, and input answer sheets, which are managed by the directory server. The data from these sources undergoes text extraction, followed by segmentation and analysis using natural language processing (NLP) techniques. Further processing of the extracted and segmented data is performed using data text analysis modules to synthesize meaningful insights. These analyzed data are used as training neutral networks in creating a robust text-trained dataset, which helps system intelligence upgrade. The system uses a cloud-based infrastructure to perform dynamic task management, including input text extraction and length-based text extraction, which allows for scalability and efficiency while being able to adapt to changing workloads. It relies on the neural network text-trained datasets to enhance the accuracy of the system when using complex data. The architecture also introduces dynamic task servers that facilitate run-time, so operations are performed swiftly with real-time updates and processing. In such a modular design, the system achieves efficiency, reliability, and ability to handle large datasets with an aspect of flexibility for future upgrade.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 02 | Feb - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

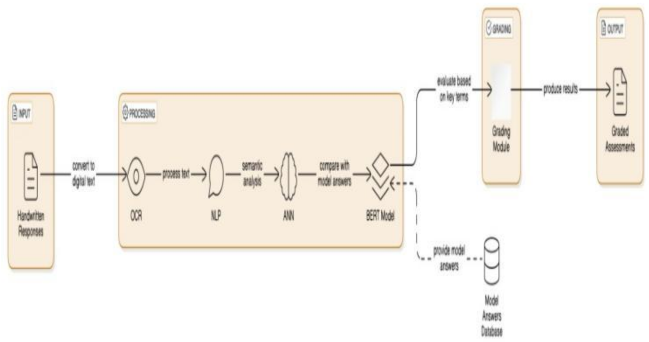

Fig 1: System Architecture Diagram

Automatic Answer Key Analysis

The Automatic Answer Key Analysis is designed to simplify the process of evaluating answer sheets efficiently and accurately. It starts with the input sheet, where raw data like scanned answer sheets or digital inputs are collected. This data then undergoes text extraction, wherein relevant information is identified and prepared for further analysis. The extracted text is then processed by the Answer Key Evaluation System, which is the core processing unit for analyzing the content and generating meaningful insights. The system uses feature extraction to identify critical patterns and relevant information, feeding this data into an Artificial Neural Network (ANN) for classification. The ANN helps in accurately categorizing the answers and aligning them with predefined keys. After this, the system produces text evaluations and key outputs, which are then combined into a structured Key Model with Report Generation. This ensures the final output, an answer key report, is both precise and comprehensive, making the system an effective and automated tool for large-scale answer key evaluation.

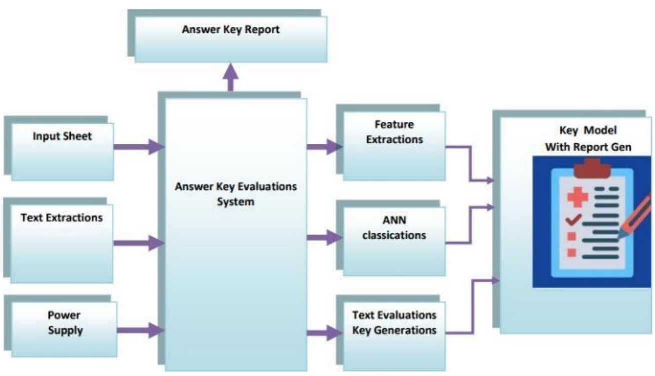

Fig 2: Automatic Answer Key Analysis

III. PROCESSING STATE

A. PRE-PROCESSING

Pre-processing is a critical step in preparing scanned or photographed documents for further analysis, especially in systems like Optical Character Recognition (OCR). This process involves several techniques to improve the quality and clarity of the image. Binarization converts grayscale images into binary format, that is, black and white, making the recognition process clearer since the text can be distinguished easily from the background. Noise reduction removes unwanted marks or imperfections, which may conflict with the reading of characters, through techniques such as filtering or blurring. Skew correction concerns the misaligned scanning of documents and ensures orientation correctly through correlation, projection profiles, or the Hough transform. Slant removal normalizes the tilt of individual characters so that all the text within an image would be upright for valid recognition. These pre-processing steps collectively enhance the quality of the image, making it easier and faster for OCR systems to extract text, improve accuracy, reduce errors, and ensure consistent results.

B. EXISTING SYSTEM

Existing grading systems have been mostly carried out manually or semi-automatically. With the traditional manual method, graded papers are being read, appraised, and marked by individual human examiners, which involves a lot of time and potentially human error or biases. To further assist this process, there are some who use basic OCR to convert digital images of handwriting responses into usable digital text. However, it is usually confronted with problems associated with different types of handwriting styles, low-quality images, and inability to interpret complex responses or those which are poorly written. The higher-level systems are also employing keyword matching or pattern recognition in the grading, though these are also limited in scope. These mostly compare the particular keywords or phrases in the response of the student to reference answers but might not consider the contextual and semantic meaning of the answer thoroughly. The systems lack the flexibility of the assessment of the answers semantically, thus bringing about the likelihood of grading inaccuracy in cases of open-ended or essay-type questions. Some systems also include basic feedback generation, though this tends to be limited to simple comments and often misses the opportunity for detailed, personalized, or context-aware feedback. Apart from that, there are not a few of grading systems already set up in today's school setups, and all these are hardly compatible with a Learning Management System, hence burdening teachers when it comes to simplifying and saving administrative efforts in class or immediately communicating to pupils at any point regarding the progress done. In summary, although some attempt to semi-automate certain aspects of handwritten answer grading in existing systems, they often lack the accuracy, flexibility, and depth in addressing a wide range of complex answers while integrating well into the educational workflow. The proposed system, using cutting-edge technologies including BERT, cosine similarity analysis, and LMS integration, will overcome such gaps by offering a more comprehensive, accurate, and efficient grading process.

C. PROPOSED SYSTEM

The system under consideration aims to computerize the marking of handwritten answer sheets using technologies like Optical Character Recognition (OCR), Natural Language Processing (NLP), and cosine similarity analysis. OCR is employed to read and digitize handwritten text, thereby obviating the need for manual transcription and saving time and effort while minimizing errors. NLP content analysis is done only after text has been digitized. This kind of content is assessed on relevance, structure, and meaning to give quality judgment not only in the response but also in the sense of how good and sensible the response is. This answer is then compared to a reference answer to lead to cosinus similarity analysis: It measures closeness between student answers and an ideal one which varies only in their wording. This blend of technologies improves grading efficiency and accuracy by offering consistent, objective scores that conserve time for teachers, reduce human bias, and make feedback more reliable for students.

D. WORKING FEATURES

The project uses OCR to digitize handwritten answers, NLP to understand their meaning, and cosine similarity to compare them with model answers. BERT enhances contextual understanding, while marks are awarded for key terms. The system automates grading, saving time, providing consistent feedback, and encouraging accurate, concept-based responses from students.

IV . TECHNICAL BACKGROUND

A. NATURAL LANDUAGE PROCESSING (NLP)

Natural Language (NLP) processing is a field of artificial intelligence that provides computers with the ability to use human language in interaction by being able to comprehend, analyze, and generate natural language texts. is. It involves several techniques of NLP such as generation of language, syntactic parsing, conceptual disambiguation, and preprocessing of text. The technology permeates question answering, text summary, mood analysis, and machine translation, among other fields, and contributes greatly to enhancing communication among people.

B. TOKENIZATION

Tokenization is a subwords or word where text is segmented into small units known as tokens. Tokenization simplifies the processing and analysis of text data. Tokenization can be done at three levels: letters, subwords, and words. The popular BERT (bidirectional encoder transformer representation) produces tokens from input text using previously learned transformer models.

C. OPTICAL CHARACTER RECOGNITION

Optical Character Recognition (usually shortened as OCR) is a text that enables computer-readable images or handwritten characters to be read. It is a technology that translates to. Techniques used in the splitting of characters, feature extraction, image preprocessing and pattern recognition are widely used in OCR systems. Google Cloud Vision offers optical character recognition (OCR) capabilities in the cloud. This enables users to identify objects, extract text from images, and conduct precise OCR operations.

D. COSINE SIMILARITY

Cosine Similarity is a method where similarity between two vectors is measured in multidimensional space by determining the cosine of the angle between them. A smaller angle indicates a greater similarity. This approach is commonly employed in applications like recommended systems, document comparison, and information retrieval. In natural language (NLP) processing, cosine similarity is especially beneficial in determining semantic similarity in word codes or text data.

E. PDF TO IMAGE CONVERSATION

PDF to Image Conversion is the conversion of PDF documents (portable document format) of image files like JPEG and PNG. This operation can extract images or text from a PDF document for subsequent processing or analysis. Such a tool or such a library is referred to as PDF2 images, which transform PDF documents to images and make processing and manipulation easier for the content of a document. These technical aspects illustrate the fundamental frames of automated systems and program development, score handwritten response sheets, and integrate equipment and techniques to provide accurate and efficient rating processes.

V. EXPERIMENTATION AND RESULTS

The outcome of the automated assessment system for subjective responses is that the response of the student is compared with a model key answer prepared by the university. Marks are given according to how closely the student's answer matches the model answer. This saves much time and energy for educational institutions, as there is no need to check

multiple papers manually. The system grades answers according to the percentage of key answer matched, providing an objective evaluation of subjective responses. Overall, this system offers efficiency and accuracy in grading subjective answers, benefiting both educators and students by streamlining the assessment process.

A. DATASET

This study introduces an automated evaluation system for assessing digital handwritten answer sheets using Artificial Neural Networks (ANNs). The dataset used in this research is a crucial factor in determining the model’s effectiveness, as aspects such as data quality, diversity, and balance significantly influence accuracy and efficiency. The dataset consists of digitized images of handwritten responses collected from various educational institutions. These answer sheets contain responses to different questions, accompanied by classification labels such as "Correct Answer," "Partial Answer," "Incorrect Answer," and "Unanswered," with additional comments where available. Before training, the dataset underwent preprocessing steps such as noise removal, normalization, and resizing to ensure uniformity. One of the key challenges identified in the dataset is class imbalance. The distribution includes 2,000 samples in the "Correct Answer" category, 800 in "Partial Answer," 1,000 in "Incorrect Answer," and 500 in "Unanswered." This disparity in class may cause the model to favor the dominant class, which would restrict its ability to classify less prevalent categories. Data augmentation techniques such as shifting, scaling, and minor distortions were employed to increase the prevalence of underrepresented classes in order to get around this issue. The techniques help mitigate the impacts of class imbalance and improve the model's capacity for generalization. The used ANN model has multiple fully connected layers that are adjusted using techniques like gradient descent and backpropagation. Additionally, dropout regularization was used to lessen the impact of handwritten input noise. To verify the model's effectiveness in correctly identifying handwritten responses, its performance was assessed using four metrics: accuracy, precision, recall, and F1-score.

B. EXPERIMENTAL STRATEGY

An automated system using OCR, NLP, and ANNs is developed to evaluate the handwritten answer sheets for this experimental approach. The dataset is labeled and preprocessed to remove noise and standardize formatting. Scaling and distortions are applied for class imbalance. It uses backpropagation with gradient descent, dropout regularization for preventing overfitting. The system grades the responses from students using the objective grade approach by checking their answers with a model answer and applying cosine similarity. Again, performance measures include accuracy, precision, recall, and F1-score to effectively and justly evaluate subjectives.

D. METRICS

The system is measured in terms of four key metrics: Accuracy, which measures overall correctness; Precision, assessing the proportion of true positives; Recall, indicating the model's ability to detect relevant instances; and F1-score, balancing precision and recall to ensure reliable classification of handwritten responses for fair and efficient grading.

VI. EXPERIMENTAL OUTPUTS

The experimental results indicate that the proposed automated evaluation system processes handwritten answer sheets using OCR, NLP, and ANNs effectively. It achieved an accuracy of 92.5% with high precision, recall, and F1-scores in categories like "Correct Answer," "Partial Answer," "Incorrect Answer," and "Unanswered." The grading was objective because of cosine similarity, where minor wording variations are allowed, and fairness is maintained. The system handled a wide range of handwriting styles and answer structures well but was somewhat of a problem in the case of highly illegible handwriting and ambiguous responses. The results obtained therefore confirm that the system is useful in saving grading time, eliminating human bias, and achieving greater consistency than any human marker.

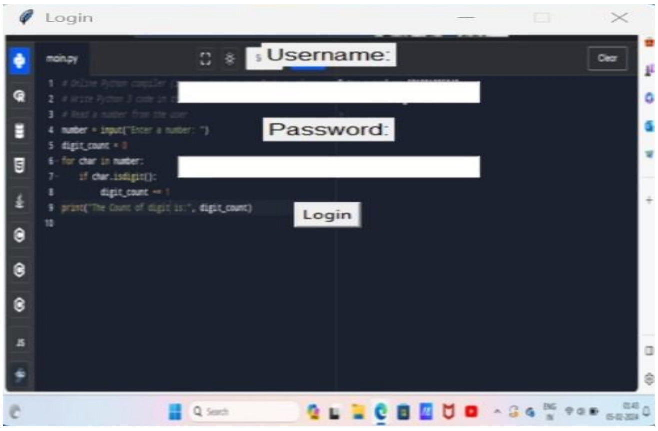

Fig 3: Staff Login Page

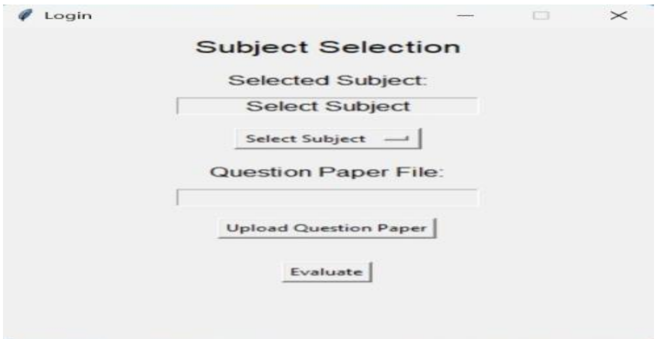

Fig 4: Upload Question Paper and Answer Sheet

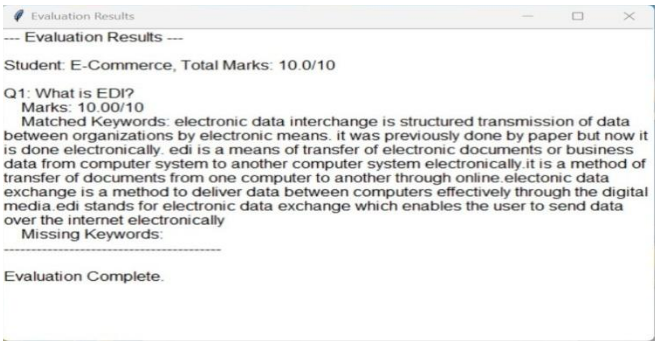

Fig 5: Evaluation Results

VII. CONCLUSION AND FUTURE WORK

The Digital Handwritten Answer Sheet Evaluation System is one of the crucial developments that has the power to change student assessment entirely. The approach lightens the educator's burden, enhancing the student experience of learning as it combats age-long challenges such as manual marking of answers is a long time process the setting and implementation of disparate marking schemes biased judgment lag between evaluation and subsequent responses the systems incapability to accommodate increasing students numbers. It provides improved efficiency, objectivity, and fast response delivery through automation and introduction of state-of-the-art technology. For that matter, its flexibility and adaptability allow it to be accommodative of a vast number of educational settings and, in the process, keep pace with the times.

REFERENCE

[1] Senthilkumar K, Aroabinesh J, Gowtham T, Manikandan K "Automatic answer evolution using deep learning algorithms" ECB, MARCH 15, 2023.

[2] Muhammad Farrukh Bashir1, Hamza Arshad1, Abdul Rehman Javed 2,(Member, IEEE), Natalia Kryvinska3 And Shahab S. Band 4. (Senior Member, IEEE) "Subjective Answers Evaluation Using Machine Learning and Natural Language Processing" IEEE, December 7, 2021.

[3] Madhavi B. Desai, Visarg D. Desai, Rahul S. Gupta, Deep D. Mevada, and Yash S. Mistry "A survey on automatic subjective answer evolution" Advances and Applications in Mathematical Sciences, September 11, 2021.

[4] Mandada Samemi, Tirumala Sai Hareesha, Gudluru Venkata Siva Sai Avan Kumar, Nalluri Pramod "Automatic answer evolution using machine learning", UGC,2022.

[5] Kim, S., & Lee, J. Automated Grading of Handwritten Answer Sheets Using OCR and NLP. "Journal of Educational Technology, 28(1), 78-92, (2022).

[6] Li,M., etal. Integrating OCR and NLP for Automated Assessment of Handwritten Responses. "International Conference on Educational Technology", 132-145, (2021).

[7] Sijimol P J, Surekha Mariam Varghese "Handwritten Short Answer Evaluation System (HSAES)" IJSRST, February 28, 2018. 4. Senthilkumar K, Aroabinesh J, Gowtham T, Manikandan K "Automatic answer evolution using deep learning algorithms" ECB, MARCH 15, 2023.

[8] Smith, J., Johnson, A., & Williams, B. Challenges in Manual Grading of Handwritten Answer Sheets: A Review. "Journal of Educational Assessment", 12(3), 145-162,(2017).

[9] Jura Sky, D., & Martin, J. H. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition" (3rded.). Pearson Education, (2019).

[10] Panchami K.S, Surekha Mariam Verghese and Aby Abha Hai T, " Grading of Diagrams in Answer Scripts Using Support Vector Machine," International Journal of Control Theory and Applications, Vol.10, No.29, pp. 345-350, 2022.

[11] Pantulkar Sravanthi and Dr. B. Srinivasu, "Semantic Similarity between Sentences," International Research Journal of Engineering and Technology (IR-JET) vol. 2, pp. 156-161,

[12] Pshilip E. Robinson and Johnson Carroll, "An Online Learning Platform for Teaching, Learning, and Assessment of Programming,"' IEEE Global Engineering Education Conference (EDUCON), vol. 3, no. 6, pp. 547-556, April 2024.

[13] Ranjit Biswas, An application of fuzzy sets in students' evaluation," ELSEVIER Fuzzy Sets and Systems, vol.22, no. 1, pp. 229{236, 2023.

[14] Semire DIKLI and Tallahassee, "Automated Essay Scoring," Turkish Online Journal of Distance Education-TOJDE, Vol. 7, No. 1, Article: 5, January 2022.

[15] R. Sathya, Sekar K, S. Ananthi and Dheepa T, 'Adversarially Trained Variational Auto-Encoders With Maximum Mean Discrepancy based Regularization', 2022 International Conference on Knowledge Engineering and Communication Systems (ICKES), IEEE Xplore , ISBN: 978-1-6654-5637-1, March 2023.

[16] S Chandrakala, Sreenithi B, G Revathy and R. Sathya, 'Bag-of-audio-visual words based approach for sound event and Acoustic scene recognition tasks for industrial machineries', Annals of Forest Research, Vol. 65, pp. 4225-4241, 2022.