VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belgaum, Karnataka – 590018

A PROJECT REPORT

ON

"AUTOMATED EVALUATION OF HANDWRITTEN ANSWER SCRIPT USING DEEP LEARNING APPROACH"

Submitted in partial fulfillment of requirements for the award of degree of Bachelor of Engineering in Information Science & Engineering during the year 2023-2024

Submitted by

NITHIN B K [4MH20IS057]

Under the Guidance
of
Prof. Siddaraj M G
Asst. Professor
Dept of IS&E
MIT Mysore.

2023 – 2024

DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING,

MAHARAJA INSTITUTE OF TECHNOLOGY MYSORE,

BELAVADI, S.R. PATNA (T), MANDYA (D) – 571477.

DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING

MAHARAJA INSTITUTE OF TECHNOLOGY MYSORE

MANDYA-571477

CERTIFICATE

Certified that the project work titled "**Automated Evaluation of Handwritten Answer Scripts Using Deep Learning Approach"** has been successfully carried out by **NITHIN B K [4MH20IS057],** is bonafide students of **Maharaja Institute of Technology, Mysore in partial fulfilment of requirements of degree of Bachelor of Engineering in Information Science & Engineering of Visveswaraya Technological University, Belgaum** during the academic year 2023-2024. It is certified that all corrections/suggestions indicated for the internal assessment have been incorporated in the report and deposited in the department library. The project report has been approved as it satisfies the academic requirements with respect to the project work prescribed for Bachelor of Engineering Degree.

Signature of guide	Signature of Principal	Signature of HOD
Siddaraj M G	**Dr. B G Naresh Kumar**	**Dr. Sharath Kumar Y H**
Asst. Professor,	Principal,	Professor & HOD,
Dept of ISE,	MIT Mysore	Dept of ISE,
MIT Mysore.		MIT Mysore.

External Viva

Name of the Examiners **Signature with date**

1.

2.

ACKNOWLEDGEMENT

It gives me immense pleasure to bring my project report entitled **“Automated Evaluation of Handwritten Answer Scripts Using Deep Learning Approach”** in the eighth semester engineering course.

I am very thankful to **Dr. B.G. Naresh Kumar**, **Principal, MITM** for supporting me in academic endeavors.

I deeply acknowledge **Dr. Sharath Kumar Y H, Associate Professor and Head of Dept. of Information Science and Engineering, MITM** for providing me timely suggestions, encouragement and support to complete this work.

I would like to sincerely thank my project guide **Siddaraj M G, Asst. Professor, Dept. of Information Science and Engineering, MITM** for providing relevant information, valuable guidance and encouragement to complete this report.

I would like to sincerely thank my project coordinator **Smithashree K P, Asst. Professor, Dept. of Information Science and Engineering, MITM** for motivating me to complete this work.

I would also like to thank all our teaching and non-teaching staff members of the department for their support.

I express my deepest gratitude and indebted thanks to **MITM** which has provided me an opportunity in fulfilling my most cherished desire of reaching the goal.

NITHIN B K [4MH20IS057]

ABSTRACT

The project focuses on the automated evaluation of handwritten answer scripts through advanced processing techniques such as OCR. By employing deep learning models, specifically Cosine similarity, Jaccard index, Eucledian distance the system effectively extracts meaningful features from handwritten text and contextual information. This approach aims to provide consistent and objective feedback across a range of subjects. Addressing several key challenges, including handwriting recognition, grading, and feedback generation, the system holds the potential to transform the assessment process. It enables educators to deliver timely and consistent evaluations, significantly reducing the manual grading burden. The system's adaptability to various educational levels and subjects enhances its versatility, making it an invaluable tool for educational institutions.

The implementation of this system not only improves efficiency in the grading process but also ensures fairness and objectivity in assessments. By leveraging the capabilities of deep learning, the system can accurately interpret and evaluate handwritten responses, providing comprehensive feedback that supports student learning and development..

Future enhancements for the automated evaluation system could involve the integration of real-time monitoring capabilities. This addition would enable the system to assess handwritten answers as they are being written, providing immediate feedback to students. Real-time monitoring can help identify and correct errors on the spot, enhancing the learning process and allowing for more interactive and engaging educational experiences. Additionally, the integration of advanced real-time processing would further automate the evaluation of handwritten answers, minimizing the need for manual intervention. This could include the use of sophisticated algorithms to detect writing patterns, recognize complex symbols, and understand context with greater accuracy. Such advancements would significantly improve the system's ability to handle a wide variety of handwriting styles and subject-specific notations, making it even more versatile and reliable.

Table of Contents

Chapter 1

INTRODUCTION

1.1 Overview

The current evaluation system for descriptive answers in academic institutions relies heavily on manual grading by faculty members. This process introduces a significant risk of variability and inconsistency in the marks awarded, as different professors might interpret and score the same answer differently. Moreover, the manual effort required to read through and evaluate numerous scripts can be overwhelming and tedious, leading to potential errors and fatigue-induced inconsistencies. These challenges highlight the need for a more efficient and reliable evaluation method that minimizes human error and ensures fair and consistent grading for all students. One of the primary issues with manual evaluation is the subjective nature of human judgment. Even with clear grading rubrics, individual interpretations and biases can lead to different scores for identical answers. Additionally, the emotional state of the evaluator at the time of grading can impact the assessment, leading to fluctuations in the quality of evaluation. These subjective elements undermine the objectivity and fairness that are crucial in academic assessments, potentially affecting students' academic records and future opportunities based on inconsistent grading standards.

To address these issues, integrating computer-based evaluation using intelligent techniques offers a promising solution. By leveraging Optical Character Recognition (OCR) and Natural Language Processing (NLP), this system can automate the evaluation of descriptive answers. OCR technology converts handwritten or printed text into machine-readable data, while NLP enables the system to understand, interpret, and analyze the natural language used by students in their answers. This approach ensures that all students' responses are evaluated using the same criteria and inference mechanisms, eliminating the variability introduced by human evaluators.

Natural Language Processing, a subfield of computer science, information engineering, and artificial intelligence, is at the heart of this automated evaluation system. NLP focuses on the interaction between computers and human languages, aiming to program computers to process and analyze large volumes of natural language data. By applying NLP techniques, the evaluation system can comprehend the context, semantics, and relevance of the answers provided by students. This technological advancement not only ensures uniformity in marking but also significantly reduces the manual effort required by faculty, allowing them to focus on more critical educational tasks. Consequently, the adoption of an NLP and OCR-based evaluation system represents a significant step forward in achieving fair, consistent, and efficient assessment in education. The ability to process and evaluate vast amounts of student scripts quickly, this system can handle large class sizes and multiple assessments simultaneously, freeing up faculty time and resources. Additionally, the data collected through such a system can provide valuable insights into common student errors and learning patterns, informing curriculum improvements and personalized feedback. As educational institutions continue to embrace digital transformation, the adoption of intelligent evaluation systems represents a pivotal shift towards more equitable and effective educational practices.

1.2 Problem Statement

The traditional manual evaluation of handwritten answer scripts in educational institutions is a time- consuming and subjective process, often leading to inconsistencies in grading. This poses challenges for educators, students, and institutions. To address these issues, there is a need for an automated system that leverages deep learning techniques to accurately and efficiently evaluate handwritten response

1.3 The Solution

Deep learning can be used to address the challenges posed by manual evaluation of handwritten scripts, we propose a comprehensive solution leveraging deep learning and associated technologies:

- Handwriting Style Adaptation: Develop deep learning models that can adapt to various handwriting styles through extensive training on diverse datasets.
- Content Analysis Models: Implement Natural Language Processing (NLP) models to assess the semantic and contextual relevance of the answers, going beyond simple text recognition

- Scalable Infrastructure: Utilize cloud computing and distributed systems to ensure scalability, enabling the processing of a large number of scripts efficiently.
- User-Friendly Feedback: Provide students with detailed, constructive feedback based on automated evaluation, enhancing the learning experience
- Robust Data Security: Prioritize data security to protect sensitive student information during the evaluation process

1.4 Existing System

In the existing system, handwritten answer scripts are manually evaluated by educators. This process involves:

- Manual Grading: Teachers or examiners assess the content, quality, and correctness of handwritten answer scripts by reading and marking them.
- Time-Consuming: Grading a large number of answer scripts can be a time-consuming process, causing delays in providing timely feedback to students.
- Resource-Intensive: In cases of large classes or high-stakes exams, a significant amount of human resources is required for grading, making it a resource-intensive process for educational institutions
- Scalability Challenges: The manual system faces scalability challenges as the volume of answer scripts increases, which can lead to inconsistencies in the grading process

1.5 Proposed System

The proposed system should overcome all the disadvantages of the existing system. The existing system is not fairness in the assessment process. Thus, the proposed system should optimized the manual evaluation and update the grading. Educators can focus on more qualitative aspects of teaching, such as providing guidance and support to students

1.6 Aim and Objectives

The aim and objectives of automated evaluation of handwritten answer scripts revolve around enhancing efficiency, accuracy, and fairness in the grading process. Here are some key aims and objectives:

- Efficiency: Automating the evaluation process aims to significantly reduce the time and effort required by human graders. This is particularly important for large-scale assessments where thousands of scripts need to be graded within a short timeframe.
- Accuracy: Automated evaluation systems strive to provide consistent and reliable grading, minimizing errors and subjective biases inherent in manual grading. By adhering to predefined criteria and algorithms, they aim to ensure fairness and reliability in the assessment process.

- Feedback: These systems can offer timely and detailed feedback to students, highlighting areas of strength and weakness in their answers. This personalized feedback can aid students in understanding their performance and areas needing improvement.

1.7 Contribution of the Study

In this project, the framework contributes to advancing technology, promoting innovation in education, improving accessibility, validating algorithms, conducting educational research, providing implementation guidelines, and driving continuous improvement in assessment practices. The project offers several key contributions:

- Consistency: By applying predefined criteria uniformly, our system ensures a consistent standard of evaluation across all answer scripts, reducing variability inherent in manual grading.
- Scalability: Our automated system can handle large volumes of answer scripts efficiently, making it suitable for use in high-stakes assessments and educational settings with large student populations.
- Feedback: Instant feedback provided by our system empowers students to identify areas of weakness and improve their understanding, enhancing the learning process.
- Data-driven Insights: Through analysis of student responses, our project generates valuable insights for educators, informing instructional strategies and curriculum design.
- Standardization: Our automated system helps maintain a standardized assessment process, ensuring fairness and equity for all students.
- Resource Optimization: By automating routine grading tasks, our project enables educators to allocate their time and expertise more effectively, focusing on activities that require critical thinking and personalized support for students.

Chapter 2

LITERATURE SURVEY

2.1 Survey Papers

[1] Text extraction using OCR: A Systematic Review

The paper, authored by Rishabh Mittal and Anchal Garg from Amity University Uttar Pradesh, explores Optical Character Recognition (OCR) technology, delving into its significance in digitizing non-digital data. It covers OCR's evolution, its applications in diverse fields like healthcare and banking, and the emergence of modern OCR systems such as Tesseract and Adobe Acrobat Pro. The review delineates the six-step process in OCR, encompassing image acquisition, preprocessing, segmentation, and morphological processing. It emphasizes the importance of each step in enhancing text recognition rates and elucidates techniques like spatial image filtering, thresholding, and morphological processing to improve OCR accuracy.

The systematic review provides a comprehensive overview of OCR technology, its applications, and the techniques used to enhance text recognition accuracy. It serves as a valuable resource for researchers, practitioners, and organizations seeking to leverage OCR for digitizing non-digital data effectively.

Beyond mere technological exploration, Mittal and Garg's work transcends into practicality by highlighting the diverse applications of OCR technology. From streamlining administrative processes in healthcare to facilitating secure transactions in banking, the authors showcase how OCR serves as a versatile solution across industries. This comprehensive overview not only underscores OCR's potential for efficiency and accuracy but also serves as a valuable resource for researchers, practitioners, and organizations seeking to harness its capabilities for digitizing non-digital data effectively. In essence, this systematic review not only encapsulates the essence of OCR technology but also serves as a beacon for understanding its transformative power in the digital age. By meticulously documenting its evolution, applications, and intricate processes, Mittal and Garg's paper offers a holistic perspective essential for anyone involved in leveraging OCR for enhanced data digitization. Its comprehensive insights and practical recommendations position it as a cornerstone resource in the realm of OCR research and implementation, empowering stakeholders to unlock new avenues of efficiency and innovation in their respective domains.

Survey findings:

- Evolution of OCR: The paper traces the evolution of OCR technology, highlighting its journey from its inception to modern advancements. It discusses how OCR has evolved from simple text recognition systems to sophisticated algorithms capable of processing complex documents.
- Applications Across Industries: The survey explores the wide-ranging applications of OCR across various industries, including healthcare and banking. It showcases how OCR technology is instrumental in automating data entry processes, improving document management, and enhancing operational efficiency in diverse sectors.

[2] Optical Character Recognition for English Handwritten Text Using Recurrent Neural Network

The paper authored by R.Parthiban, R.Ezhilarasi, and D.Saravanan from IFET College of Engineering explores an Optical Character Recognition (OCR) system using Recurrent Neural Networks (RNN) for handwritten English text. Highlighting the challenges in accurately recognizing handwritten content due to varied styles, it employs a bottom-up methodology for enhancement. The proposed model utilizes RNN layers followed by Connectionist Temporal Classification (CTC) to improve accuracy. Divided into four main stages—Dataset collection, Pre-processing, Training and Testing Set division, and RNN implementation—it achieves successful recognition of handwritten characters and digits. Python and TensorFlow are used for system implementation. The paper concludes by achieving a 90% accuracy in recognizing handwritten English documents and suggests further research for enhancing the system's performance**.**

The paper meticulously outlines the four main stages of their methodology: Dataset collection, Pre-processing, Training and Testing Set division, and RNN implementation. Each stage is crucial in the development and refinement of the OCR system, highlighting the authors' comprehensive approach to tackling the intricacies of handwritten text recognition. Furthermore, the choice of Python and TensorFlow for system implementation underscores the emphasis on leveraging state-of-the-art tools and technologies for optimal performance. A notable achievement of the study is the reported 90% accuracy in recognizing handwritten English documents. This milestone underscores the efficacy of the proposed RNN-based OCR system in addressing the challenges posed by handwritten content. However, the authors also acknowledge the potential for further research to enhance the system's performance. This recognition of ongoing improvement reflects a commitment to continuous innovation and refinement in the field of OCR, paving the way for even more accurate and robust solutions in the future.

Survey findings:

- Modern OCR Systems: The review identifies and discusses modern OCR systems such as Tesseract and Adobe Acrobat Pro. It provides insights into their features, capabilities, and areas of application, demonstrating how these systems have revolutionized text extraction and document processing.
- Six-Step Process in OCR: The paper delineates the six-step process involved in OCR, including image acquisition, preprocessing, segmentation, and morphological processing. It emphasizes the importance of each step in achieving high text recognition rates and discusses techniques such as spatial image filtering, thresholding, and morphological processing to enhance OCR accuracy.

[3] Handwritten Text Recognition using Deep Learning

The paper authored by Nikitha A explores Handwritten Text Recognition Using Deep Learning, Nikitha A's paper delves into Handwritten Text Recognition (HTR) using Deep Learning algorithms, aiming to improve accuracy in transforming handwritten texts into digital forms. Employing LSTM-based models, the study emphasizes recognizing words rather than characters, achieving notable accuracy enhancements. The paper explores the integration of HTR into OCR systems, demonstrating significant improvements with a 2DLSTM approach compared to conventional methods. Addressing challenges in training datasets, it draws insights from various sources including IAM datasets and online/offline handwritten samples. Through pre-processing techniques like noise removal, segmentation, and binarization, the study optimizes image processing for accurate recognition. The proposed methodology showcases promising results, achieving a 94% accuracy rate in recognizing handwritten text. Additionally, the comparison table highlights the superiority of the LSTM-based model in character and word recognition rates. The paper concludes by suggesting scalability for multilingual evaluations and underscores the potential for future advancements in handwritten text recognition systems.

The methodology yields promising results, with the reported 94% accuracy rate in recognizing handwritten text serving as a testament to its effectiveness. Moreover, the comparison table presented in the paper underscores the superiority of LSTM-based models in both character and word recognition rates, further validating the efficacy of the approach. However, the author doesn't rest on these achievements alone; instead, the paper concludes by suggesting avenues for scalability, particularly in multilingual evaluations. This forward-thinking perspective highlights the author's awareness of the evolving landscape of handwritten text recognition and underscores the potential for future advancements in the field.

Survey findings:

- Deep Learning Approach: The study employs LSTM-based models for HTR, emphasizing word recognition over character recognition. This approach leads to notable accuracy enhancements compared to traditional methods.
- Integration with OCR Systems: The paper explores integrating HTR into OCR systems, demonstrating significant improvements with a 2DLSTM approach compared to conventional methods. This integration enhances the overall performance of OCR systems for handwritten text recognition tasks.

[4] Extraction of Information from Handwriting using Optical Character recognition and Neural Networks

Piyush Mishra, from Sardar Patel Institute Of Technology, explores Handwritten Text Recognition using Optical Character Recognition (OCR) and Neural Networks. The paper delves into image processing techniques to detect handwritten text, converting it into a digital format via TensorFlow. By merging Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), the study aims to bridge the gap between paper and digital worlds. They use a comprehensive dataset, IAM, to train the model, achieving 11.29% average character error rate (CER) and 37.34% word error rate (WER). The paper suggests enhancing model accuracy with larger datasets, mobile app development, and customizing applications for specific user needs like cheque forgery detection or aiding the visually impaired.

The paper suggests avenues for further improving model accuracy and expanding the applicability of the HTR system. One such recommendation is the utilization of larger datasets, which can help fine-tune the model and address potential limitations. Additionally, the proposal for mobile app development signifies the potential for democratizing access to HTR technology, allowing users to seamlessly digitize handwritten text on-the-go. Moreover, the customization of applications for specific user needs, such as cheque forgery detection or assisting the visually impaired, underscores the versatility and societal impact of Mishra's research.In essence, Piyush Mishra's work not only contributes to the advancement of HTR technology but also underscores its potential to address real-world challenges and cater to diverse user needs. By blending cutting-edge techniques from OCR and Neural Networks with practical applications and future-oriented recommendations, Mishra's paper serves as a blueprint for leveraging technology to bridge the gap between analog and digital realms.

Survey findings:

- Training Dataset Challenges: Addressing challenges in training datasets, the study draws insights from various sources, including IAM datasets and online/offline handwritten samples. This comprehensive approach ensures robust model training and generalization.
- Pre-processing Techniques: Through pre-processing techniques such as noise removal, segmentation, and binarization, the study optimizes image processing for accurate recognition. These techniques improve the quality of input data, leading to better performance of the HTR models.

[5] Natural Language Processing Journal

Prabakaran N., Kannadasan R., Krishnamoorthy A., and Vijay Kakani from Vellore Institute of Technology and Inha University present an innovative method in Natural Language Processing Journal. Their Bidirectional LSTM approach focuses on auto-evaluating written scripts using deep learning models, aiming to mimic human scoring for English characters and numbers. The system relies on Handwritten Text Recognition (HTR) to process custom keywords provided by invigilators for checking answers. It aims to reduce manual checking by employing BiLSTM and CRNN, achieving better text recognition and error identification, benefitting both students and teachers in terms of time efficiency. The model integrates datasets like IAM and Washington to improve character recognition, presenting an approach that aligns with the transitioning trend towards online exams featuring handwritten responses. This approach facilitates the transformation of scanned handwritten copies into digital assessments, streamlining evaluation processes.

The model's reliance on datasets like IAM and Washington underscores its robustness and adaptability, as it leverages diverse sources to improve character recognition. This integration of multiple datasets demonstrates the authors' commitment to developing a versatile solution that can handle various handwriting styles and contexts. Moreover, the approach aligns with the transitioning trend towards online exams featuring handwritten responses, facilitating the seamless transformation of scanned handwritten copies into digital assessments. Overall, the research represents a significant step forward in automating the evaluation of written scripts through deep learning techniques. By combining innovative methodologies with practical applications, their paper not only addresses the challenges associated with manual grading but also contributes to streamlining evaluation processes in educational settings. This approach holds promise for revolutionizing assessment practices and enhancing the efficiency and accuracy of handwritten text evaluation.

Survey findings:

- Performance Evaluation: The proposed methodology showcases promising results, achieving a 94% accuracy rate in recognizing handwritten text. Additionally, the comparison table highlights the superiority of the LSTM-based model in character and word recognition rates compared to traditional methods.
- Scalability and Future Directions: The paper suggests scalability for multilingual evaluations, indicating the potential for adapting the HTR system to recognize handwritten text in various languages. It also underscores the potential for future advancements in handwritten text recognition systems, suggesting avenues for further research and development.

[6] Expert Systems With Applications

Faisal Jamil and Ibrahim A. Hameed propose an intelligent open-ended question evaluation model focusing on students' answers. They highlight the limitations of locked trained models in adapting to dynamic changes, aiming to enhance performance through predictive optimization using deep neural networks. Their approach considers question types, structure, keywords, language, and conceptual aspects, employing semantic similarity scores derived from the WordNet library and Growbag dataset. They aim to create an automated grading system using natural language processing techniques and supervised machine learning classification. The research aims to address the challenges in evaluating open-ended questions, enhancing fairness, and providing consistent assessments. They incorporate predictive optimization, semantic similarity algorithms, and machine learning to optimize grading accuracy in a smart learning environment. The study's extensive experiments and ablation studies underscore the significance of their proposed mechanism, focusing on the practicality and performance of their intelligent evaluation system for open-ended questions.

The research underscores the importance of fairness and consistency in assessments by incorporating predictive optimization, semantic similarity algorithms, and machine learning into the grading system. By optimizing grading accuracy in a smart learning environment, Jamil and

Hameed's study aims to provide educators with a reliable tool for assessing open-ended responses effectively. Moreover, the extensive experiments and ablation studies conducted by the authors validate the practicality and performance of their proposed mechanism, highlighting its potential to significantly enhance the evaluation process for open-ended questions in educational settings.

Survey findings:

- Limitations of Locked Trained Models: The authors highlight the limitations of locked trained models in adapting to dynamic changes in student responses. They emphasize the need for models that can enhance performance through predictive optimization using deep neural networks.
- Consideration of Question Characteristics: Their approach considers various aspects of open-ended questions, including question types, structure, keywords, language, and conceptual aspects. By taking into account these characteristics, they aim to develop a more robust evaluation model.

[7] Knowledge-Based Systems

Xia Li, Minping Chen, and Jian-Yun Nie introduce SEDNN, a novel approach in their paper "Shared and Enhanced Deep Neural Network Model for Cross-Prompt Automated Essay Scoring." They tackle the challenge of cross-prompt AES, aiming to extract broader, transferable knowledge from multiple source prompts to enhance performance in rating essays for a target prompt. Their model, SEDNN, diverges from conventional approaches by amalgamating data from various source prompts to extract both prompt-independent and prompt-dependent features. Instead of solely focusing on invariant prompts, they contrast the target prompt with the entirety of source prompts, aiming to capture a richer set of transferable knowledge, including general language features and prompt-specific attributes.

SEDNN comprises two components: a shared model (SModel) and an enhanced model (EModel). SModel is trained to extract shared knowledge across prompts, incorporating prompt-independent and potentially prompt-dependent features. This shared knowledge aids in generating pseudo training data for the target prompt. EModel, a Siamese network-based framework, is then trained to integrate more prompt-specific features from the target prompt essays and title, refining the scoring accuracy further. Their innovative approach demonstrates superior performance in cross-prompt AES, outperforming existing methods. By leveraging a comprehensive set of shared and prompt-specific features, SEDNN showcases the potential to revolutionize automated essay scoring by effectively transferring knowledge across prompts for enhanced evaluation accuracy.

Survey findings:

- Semantic Similarity Scores: The authors leverage semantic similarity scores derived from the dataset to assess the similarity between student responses and model answers. This approach enhances the accuracy of evaluating open-ended questions by considering the semantic relevance of student answers.
- Automated Grading System: The research aims to create an automated grading system using natural language processing techniques and supervised machine learning classification. By automating the grading process, they seek to address the challenges in evaluating open-ended questions while ensuring fairness and consistency in assessments.

[8] Unconstrained Offline Handwritten Word Recognition by Position Embedding Integrated ResNets Model

The paper authored by Xiangping Wu, Qingcai Chen, Jinghan You, and Yulun Xiao explores Unconstrained Offline Handwritten Word Recognition by Position Embedding Integrated ResNets Model.The paper introduces a new method for offline handwritten word recognition, merging position embeddings with ResNets and BiLSTM networks. It avoids complex linguistic knowledge, achieving top results on datasets like the 2017 ICDAR competition and RIMES public dataset. The model combines ResNets to extract image features and position embeddings to represent character sequences. This approach simplifies segmentation-free word recognition, making it adaptable to resource-limited languages. By incorporating these elements, the proposed model performs impressively on various handwriting recognition benchmarks, showcasing its potential for multilingual applications.

The key highlights of the proposed model is its impressive performance across various handwriting recognition benchmarks. By incorporating position embeddings into ResNets and BiLSTM networks, the model achieves state-of-the-art results, underscoring its potential for multilingual applications. This versatility makes the model well-suited for diverse linguistic contexts, where handwriting styles and conventions may vary widely. As a result, the proposed approach opens up new possibilities for advancing handwritten word recognition technology and addressing the needs of multilingual communities. In summary, the paper presents a groundbreaking method for offline handwritten word recognition that combines ResNets, position embeddings, and BiLSTM networks. This integration of cutting-edge techniques not only simplifies the recognition process but also enhances the model's adaptability and performance across different languages and datasets. The proposed approach holds promise for advancing handwriting recognition technology and facilitating its adoption in diverse real-world applications.

Survey findings:

- Method Overview: The paper introduces a novel method for offline handwritten word recognition that integrates position embeddings with ResNets (Residual Networks) and Bidirectional Long Short-Term Memory (BiLSTM) networks.
- Simplified Approach: Unlike traditional methods that rely on complex linguistic knowledge or explicit segmentation, the proposed model simplifies word recognition by utilizing ResNets to extract image features and incorporating position embeddings to represent character sequences. This segmentation-free approach streamlines the recognition process, making it adaptable to languages with limited linguistic resources.

[9] Automated Language Scoring System by Employing Neural Network Approaches.

Authors: Anne Kwong, Junaid Hussain Muzamal, Usman Ghani Khan.The research delved into enhancing automated language scoring using a sophisticated blend of neural network architectures, particularly attention-based BLSTM-RNNs. By dissecting language proficiency into dimensions like delivery, grammar, and content, the proposed methodology showcased remarkable improvements in correlation with human-rated scores across monologues and dialogues.

The proposed methodology showcased remarkable improvements in correlation with human-rated scores across both monologues and dialogues. By leveraging attention mechanisms within BLSTM-RNNs, the model exhibits enhanced capabilities in capturing relevant linguistic features while assigning scores. This attention-based approach enables the model to dynamically focus on salient aspects of language proficiency, such as coherence, fluency, and lexical richness, thereby mimicking the nuanced judgment process of human raters more effectively.

The incorporation of attention mechanisms within BLSTM-RNNs contributes to the model's interpretability, allowing for insights into which linguistic features are deemed most significant in determining language proficiency scores. This transparency enhances the trustworthiness of automated scoring systems, as stakeholders can gain a clearer understanding of the factors driving the assigned scores.The research underscores the potential of advanced neural network architectures, particularly attention-based BLSTM-RNNs, in enhancing automated language scoring systems. By adopting a multidimensional approach to language proficiency evaluation and leveraging attention mechanisms, their methodology demonstrates significant strides in aligning automated scoring with human-rated assessments. This has profound implications for various domains, including education, language testing, and automated speech recognition, where accurate and reliable language proficiency assessment is paramount.

Survey findings:

- Performance: The model achieves state-of-the-art results on benchmark datasets such as the 2017 ICDAR competition and the RIMES public dataset. This demonstrates its effectiveness in recognizing unconstrained offline handwritten words.
- Multilingual Applications: Due to its simplified architecture and reliance on image features and position embeddings, the proposed model shows promise for multilingual applications. It can be adapted to recognize handwriting in various languages without requiring extensive linguistic resources or language-specific knowledge.

[10] A set of benchmarks for Handwritten Text Recognition on historical documents.

The paper provides a comprehensive overview of the advancements in Off-line Handwritten Text Recognition (HTR), emphasizing its crucial role in unlocking vast historical manuscripts for accessibility and searchability. It explores the challenges shared between handwriting and speech processing, drawing parallels between the two and highlighting the successes achieved by employing techniques from Automatic Speech Recognition (ASR) in HTR. The article also presents various publicly available datasets used in HTR, detailing their characteristics and performance benchmarks, as well as the impact of European projects like transcriptorium and READ in advancing HTR technologies through competitions and collaborations with archives and libraries.

The paper underscores the importance of publicly available datasets in driving HTR research forward. By detailing the characteristics and performance benchmarks of various datasets used in HTR, the paper provides valuable resources for evaluating and benchmarking HTR algorithms. Additionally, the paper highlights the significant impact of European projects such as transcriptorium and READ in fostering HTR advancements through collaborative initiatives and competitions. These projects serve as catalysts for innovation, facilitating knowledge exchange and technological progress in the field of HTR. The paper offers a comprehensive exploration of Off-line Handwritten Text Recognition, from its fundamental challenges to its transformative potential. By elucidating the interplay between handwriting and speech processing, detailing publicly available datasets, and highlighting the impact of collaborative initiatives, the paper provides valuable insights into the past, present, and future of HTR technology.

Survey findings:

- Interplay between Handwriting and Speech Processing:The paper highlights the influence of Automatic Speech Recognition (ASR) techniques on Off-line Handwritten Text Recognition (HTR), emphasizing shared challenges and insights that have led to advancements in HTR technology.
- Impact of Publicly Available Datasets and European Projects:The survey underscores the significance of publicly available datasets for evaluating HTR algorithms and acknowledges the substantial contributions of European projects like transcriptorium and READ in driving HTR advancements through collaborations and competitions.

[11] HTR-Flor: A Deep Learning System for Offline Handwritten Text Recognition.

The paper introduces a novel Gated Convolutional Recurrent Neural Network (Gated-CRNN) architecture for Handwritten Text Recognition (HTR) systems. It addresses the limitations of existing models by proposing a compact yet effective approach that outperforms state-of-the-art architectures by 33% across five benchmark datasets. The Gated-CRNN model exhibits strong recognition capabilities, even with limited training data, and requires fewer parameters, making it computationally efficient and suitable for real-world applications like smartphones and robots.

The success of the Gated-CRNN model is its unique architectural design, which seamlessly integrates gated convolutional and recurrent layers. This combination allows the model to capture both spatial and sequential dependencies inherent in handwritten text, thereby enhancing recognition capabilities. Moreover, the compact nature of the Gated-CRNN architecture enables it to achieve superior performance while requiring fewer parameters compared to existing models. This not only enhances computational efficiency but also makes the model more accessible and practical for deployment in real-world applications.

The key strengths of the Gated-CRNN model is its ability to generalize well even with limited training data. This adaptability is crucial for scenarios where acquiring large annotated datasets may be challenging or impractical. Despite the scarcity of training data, the Gated-CRNN demonstrates strong recognition capabilities, underscoring its versatility and suitability for various HTR tasks.The inherent efficiency and effectiveness of the Gated-CRNN architecture make it an ideal candidate for deployment in resource-constrained environments, such as smartphones and robots. Its lightweight design and superior performance position the Gated-CRNN as a practical solution for real-world applications, where computational resources may be limited. Overall, the paper's findings not only showcase the advancements in HTR technology but also underscore the transformative potential of the Gated-CRNN architecture in enhancing the accessibility and efficiency of handwritten text recognition systems.

Survey findings:

- The paper demonstrates that the proposed Gated Convolutional Recurrent Neural Network (Gated-CRNN) architecture surpasses state-of-the-art models by 33% across five benchmark datasets for Handwritten Text Recognition (HTR). This finding highlights the effectiveness and robustness of the Gated-CRNN model in accurately recognizing handwritten text, showcasing its potential to advance the field of HTR significantly.
- The study reveals that the Gated-CRNN model exhibits strong recognition capabilities even with limited training data while requiring fewer parameters compared to existing architectures. This compact yet effective design makes the Gated-CRNN model computationally efficient, rendering it suitable for real-world applications such as smartphones and robots.

[12] Intelligent handwritten recognition using hybrid CNN architectures based-SVM classifier with dropout.

The proposed research presents a sophisticated approach utilizing CNN-based-SVM with dropout for Arabic handwriting recognition, achieving impressive accuracy rates across various datasets. The study systematically compares different classifiers, loss functions, and execution times, demonstrating superior performance with CNN-based-SVM. The model outperforms prior state-of-the-art methods on multiple datasets, showcasing its robustness and efficiency in document image classification. Complexity: The proposed model's complexity might hinder its widespread adoption, requiring substantial computational resources.Deep neural networks like CNNs can lack interpretability, making it challenging to understand how specific decisions are made, posing a challenge in explaining the model's reasoning in practical applications.

The CNN-based-SVM model showcases robustness and efficiency in Arabic handwriting recognition tasks, its complexity necessitates careful consideration of deployment scenarios and computational resources. Practical implementation of the proposed approach may require significant investments in computational infrastructure and expertise, limiting its accessibility to certain organizations or individuals. Therefore, addressing the complexity of the model and optimizing its computational efficiency are critical steps towards ensuring broader adoption and practical utility in real-world settings.

The challenges posed by its complexity, the CNN-based-SVM model represents a significant advancement in Arabic handwriting recognition technology. Its superior performance and accuracy rates underscore its potential to revolutionize document image classification tasks, offering valuable insights into the capabilities of hybrid neural network architectures. Moving forward, efforts to streamline the model's complexity and enhance its computational efficiency will be essential in unlocking its full potential and facilitating its integration into practical applications and workflows.

Survey findings:

- Resource Efficiency: The integration of ResNets and position embeddings enhances the efficiency of the recognition process, making it suitable for resource-limited environments or applications where computational resources are constrained.
- Generalizability: The model's impressive performance across multiple benchmark datasets highlights its generalizability and robustness in handling diverse handwriting styles and conditions.

[13] TextCaps : Handwritten Character Recognition with Very Small Datasets

The paper titled "TextCaps: Handwritten Character Recognition with Very Small Datasets" authored by Vinoj Jayasundara, Sandaru Jayasekara, Hirunima Jayasekara, Jathushan Rajasegaran, Suranga Seneviratne, and Ranga Rodrigo addresses the challenge of limited labeled training data for localized languages in handwritten character recognition. The authors propose a novel approach using Capsule Networks (CapsNets) to generate new training samples with realistic augmentations, overcoming the limitations of deep learning techniques with small datasets. The technique involves manipulating instantiation parameters and adding controlled noise to reflect variations in human handwriting. The results demonstrate surpassing state-of-the-art performance on the EMNIST-letter dataset with just 200 training samples per class. The paper also explores the application of this approach in other datasets, such as EMNIST-balanced, EMNIST-digits, and MNIST, showcasing its effectiveness in character recognition for languages with limited labeled data. The authors highlight the contributions of their work, including outperforming existing results, evaluating flexibility on non-character datasets like Fashion-MNIST, and proposing a strategy for effective use of loss functions to improve reconstructions.

The results demonstrate the efficacy of the proposed approach, with the model surpassing state-of-the-art performance on the EMNIST-letter dataset using just 200 training samples per class. Furthermore, the paper explores the application of TextCaps in other datasets such as EMNIST-balanced, EMNIST-digits, and MNIST, showcasing its versatility and effectiveness in character recognition tasks for languages with limited labeled data. This underscores the potential of the TextCaps approach to address the challenges posed by data scarcity in various domains.

The authors highlight several significant contributions of their work, including the outperformance of existing results, the evaluation of flexibility on non-character datasets like Fashion-MNIST, and the proposal of a strategy for effectively utilizing loss functions to improve reconstructions. These contributions not only advance the field of handwritten character recognition but also pave the way for more robust and efficient models capable of handling data scarcity and achieving superior performance across diverse datasets and applications.

Survey findings:

- The paper's findings underscore the effectiveness of Capsule Networks (CapsNets) in generating new training samples with realistic augmentations, thereby overcoming the challenges posed by limited labeled training data in handwritten character recognition tasks. This approach not only enhances model performance but also facilitates the generalization of models to unseen data, making it a promising solution for languages with sparse labeled datasets.
- The survey reveals that the TextCaps approach demonstrates superior performance not only on character recognition tasks but also on non-character datasets such as Fashion-MNIST. This highlights the versatility and effectiveness of the proposed method in various applications beyond handwritten character recognition, showcasing its potential for broader use cases and domains.

[14] Hybrid CNN-SVM Classifier for Handwritten Digit Recognition

The paper titled "Hybrid CNN-SVM Classifier for Handwritten Digit Recognition" by Savita Ahlawat and Amit Choudhary proposes a hybrid model combining Convolutional Neural Networks (CNN) and Support Vector Machine (SVM) for handwritten digit recognition using the MNIST dataset. In this hybrid approach, CNN serves as an automatic feature extractor, while SVM functions as a binary classifier. The experimental results demonstrate the effectiveness of the proposed framework, achieving a high recognition accuracy of 99.28% over the MNIST handwritten digits dataset. The paper highlights the synergy between CNN's feature extraction capabilities and SVM's classification power, addressing the challenges of diverse and distorted handwritten digit images. Manual handcrafted feature extraction methods are deemed tedious and may compromise efficiency versus recognition accuracy trade-offs. The proposed hybrid CNN-SVM approach may face challenges in handling complex variations in handwriting patterns, potentially affecting its performance on diverse datasets.

The paper demonstrate the effectiveness of the hybrid CNN-SVM model, achieving an impressive recognition accuracy of 99.28% over the MNIST handwritten digits dataset. This high accuracy underscores the synergy between CNN's feature extraction capabilities and SVM's classification power, emphasizing the complementary nature of the two methods in handling diverse and distorted handwritten digit images. Moreover, the hybrid approach eliminates the need for manual handcrafted feature extraction methods, which are often tedious and may compromise the efficiency versus recognition accuracy trade-offs.

The proposed hybrid CNN-SVM approach may encounter challenges in handling complex variations in handwriting patterns. Handwritten digits can exhibit diverse styles, sizes, and distortions, which may pose difficulties for the model in accurately categorizing them. The inherent variability in handwriting across different individuals adds another layer of complexity, potentially affecting the model's performance on diverse datasets beyond MNIST. Thus, while the hybrid CNN-SVM model demonstrates promising results on the MNIST dataset, its robustness and generalizability to real-world handwritten digit recognition scenarios warrant further investigation and evaluation.

Survey findings:

- The survey findings highlight the strengths of hybrid CNN-SVM models in combining automatic feature extraction with powerful classification capabilities for handwritten digit recognition tasks. This approach leverages the complementary nature of CNN and SVM, allowing for efficient processing of complex image data while achieving high recognition accuracy.

- The survey underscores the challenges faced by hybrid CNN-SVM models in handling complex variations in handwriting patterns, especially when confronted with diverse styles, sizes, and distortions in handwritten digits. Addressing these challenges is crucial for enhancing the robustness and generalizability of the model across diverse datasets and real-world scenarios.

[15] A Study of Automated Evaluation of Student's Examination Paper using Machine Learning.

In their collaborative research, Ganga Sanuvala, a Research Scholar, and Syeda Sameen Fatima, a retired Professor from the Department of CSE at Osmania University, present a groundbreaking solution titled "Handwritten Answer Evaluation System Using OCR and ML/NLP Techniques." The study addresses the challenges of traditional evaluation processes in education, notorious for their high human effort requirements and time-consuming nature. Through the implementation of Optical Character Recognition (OCR) and advanced Machine Learning/Natural Language Processing (ML/NLP) techniques, the proposed Handwritten Answer Evaluation System (HAES) offers an automated mechanism for text extraction and grading from handwritten answer sheets.

HAES stands out as an innovative response to the growing need for efficient evaluation systems in the academic landscape. By leveraging OCR for text extraction and ML/NLP for grading based on cosine set similarity measures, the system provides a reliable and timely alternative to labor-intensive evaluation methods. The authors emphasize the potential benefits of HAES, including enhanced efficiency, reduced evaluation time, and improved resource utilization. Their research aims to revolutionize the assessment process, offering a promising solution to the challenges posed by traditional handwritten exam evaluations.

Survey findings:

- Efficiency Boost: Survey respondents noted a significant improvement in evaluation efficiency with the implementation of HAES, with a majority reporting reduced time spent on grading handwritten answer sheets compared to traditional manual methods.
- Accuracy Enhancement: HAES demonstrated higher accuracy levels in grading compared to human evaluators, particularly in cases involving large volumes of answer sheets. This

accuracy improvement was attributed to the consistent application of ML/NLP techniques for grading.

[16] An Automated Essay Scoring Systems: A Systematic Literature Review

The paper explores the significance of assessment in the education system, emphasizing the challenges posed by the manual evaluation process, such as time consumption and lack of reliability. It highlights the evolution of online examination systems as an alternative tool, particularly for essays and short answers. The systematic literature review delves into Artificial Intelligence and Machine Learning techniques employed in automated essay scoring systems, addressing the limitations of current studies and research trends. The authors discuss the shift to an online educational system, the role of AES in computer-based exams, and the crucial parameters for evaluating essays, including relevance, coherence, and cohesion. The review covers the historical development of AES, spanning from traditional approaches to recent advancements, such as deep learning techniques. The challenges and limitations in current research, along with the need for comprehensive evaluation criteria, are underscored. The paper concludes by proposing avenues for future research in automated essay scoring.

The paper explores the application of Artificial Intelligence (AI) and Machine Learning (ML) techniques in automated essay scoring (AES) systems. It provides insights into the current state of research, addressing both the strengths and limitations of existing studies while also identifying emerging trends in the field. By examining the evolution of AES, from traditional approaches to contemporary deep learning techniques, the paper offers a comprehensive overview of the historical development of automated essay scoring technology. The authors delve into the crucial parameters for evaluating essays, including relevance, coherence, and cohesion. This discussion underscores the multifaceted nature of essay assessment and emphasizes the need for comprehensive evaluation criteria in AES systems. By considering these key parameters, AES technologies can more effectively mimic human evaluative processes, thereby enhancing the accuracy and reliability of automated essay scoring.

The advancements in AES technology, the paper also highlights several challenges and limitations present in current research. These include issues related to model generalizability, bias detection, and the incorporation of nuanced language understanding. Addressing these challenges will be essential for advancing the field of automated essay scoring and ensuring its broader applicability in educational contexts. In conclusion, the paper proposes avenues for future research, emphasizing the importance of continued innovation and collaboration in the development of more robust and effective AES systems.

Survey findings:

- Educator Satisfaction: Educators expressed satisfaction with HAES, citing its user-friendly interface, intuitive operation, and ability to streamline evaluation workflows. Many reported increased job satisfaction due to reduced administrative burdens associated with manual grading.
- Student Performance Insights: HAES provided valuable insights into student performance trends and learning outcomes through its analytics and reporting features. Educators noted the ability to identify common misconceptions, areas of strength, and opportunities for intervention based on automated grading results.

[17] UESTS: An Unsupervised Ensemble Semantic Textual Similarity Method

The paper introduces UESTS, an Unsupervised Ensemble Semantic Textual Similarity method, addressing the task of assessing similarity between two texts based on meaning. Key contributions include a novel synset-oriented word aligner using BabelNet, three unsupervised STS approaches, and an ensemble method (UESTS) that combines various similarity measures. The proposed methods outperform existing unsupervised approaches, providing a promising solution for semantic similarity assessment without the need for training data.

The key strengths of UESTS lies in its ensemble method, which integrates the outputs of the three unsupervised STS approaches to produce a final similarity score. This ensemble strategy leverages the diversity of the individual similarity measures, mitigating the weaknesses of any single approach and yielding more reliable similarity assessments. As a result, UESTS consistently outperforms existing unsupervised methods, demonstrating its effectiveness in accurately capturing semantic similarity between texts.

The introduction of UESTS represents a significant advancement in the field of semantic textual similarity assessment. By combining innovative techniques such as synset-oriented word alignment, unsupervised STS approaches, and ensemble methodology, UESTS offers a promising solution for evaluating semantic similarity without the need for annotated training data. This approach holds considerable potential for various applications, including information retrieval, natural language processing, and automated text analysis.

Survey findings:

- Resource Optimization: The implementation of HAES resulted in optimal resource utilization within educational institutions. By automating the evaluation process, institutions were able to reallocate human resources to more strategic tasks such as curriculum development and student support services.
- Scalability and Adaptability: HAES demonstrated scalability and adaptability across different educational contexts and subject areas. Survey respondents highlighted its ability to accommodate various question formats, languages, and grading criteria, making it suitable for diverse academic settings.

Chapter 3

SOFTWARE REQUIREMENT SPECIFICATION

3.1 System requirements

3.1.1 Stakeholders:

- Team Members
- Project Guide
- Students
- Project Reviewers
- Faculty Department
- College Management
- Organization’s Officials
- Admin

3.1.2 Functional requirements:

- Image Preprocessing: The system should be able to preprocess scanned or photographed handwritten answer scripts to enhance readability and remove noise. Preprocessing techniques may include resizing, normalization, noise reduction, and contrast adjustment.
- Handwriting Recognition: The system should accurately recognize handwritten text from answer scripts using deep learning models such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs).It should support the recognition of diverse handwriting styles, languages, and characters.
- Answer Extraction: The system should extract individual answers from the scanned or photographed answer scripts. It should handle various answer formats, including single-line, multi-line, and paragraph-style answers.
- Answer Grading: The system should employ deep learning techniques for grading extracted answers based on predefined criteria or rubrics. It should support the assessment of both quantitative and qualitative responses, including numerical calculations and essay-style answers.
- Feedback Generation: The system should generate detailed feedback for each graded answer, highlighting strengths, weaknesses, and areas for improvement. Feedback may include scores, comments, suggestions, and references to relevant course materials. Integration with Learning

- Management Systems (LMS):The system should seamlessly integrate with existing Learning Management Systems (LMS) used by educational institutions. Integration should support features such as result synchronization, gradebook updates, and student feedback delivery.

3.1.3 Non functional requirements:

- Accuracy: The system should achieve high accuracy in handwriting recognition and answer grading to ensure reliable assessment results. Accuracy requirements may vary depending on the specific evaluation criteria and educational standards.
- Scalability: The system should be scalable to handle large volumes of answer scripts during peak evaluation periods, such as exams or assignments. Scalability should ensure consistent performance without degradation in speed or accuracy.
- Performance: The system should deliver fast response times for processing and grading handwritten answer scripts. Performance requirements should consider factors such as system latency, throughput, and responsiveness.
- Security: The system should ensure the confidentiality and integrity of student data and evaluation results. Security measures should include user authentication, access control, data encryption, and audit trails to prevent unauthorized access or tampering.
- Usability: The system should have an intuitive user interface that is easy to navigate for both educators and administrators. Usability requirements should prioritize user experience, accessibility, and support for multiple devices and screen sizes.
- Robustness: The system should be robust against variations in handwriting styles, image quality, and environmental factors. Robustness requirements should include robust deep learning models, error handling mechanisms, and quality assurance procedures.
- Reliability and Availability: The system should demonstrate high reliability and availability to ensure uninterrupted operation during critical evaluation periods, such as exam seasons.

3.1.4 Hardware requirements:

- System Processor - Intel core i5 or above.
- Hard disk - 500GB or more.
- RAM - 4GB or more.
- Monitor - 15 VGA colour or above.
- GPU-Nvidia GeForce 940MX

3.1.5 Software requirements:

- Operating System - Windows XP/7/8/10.
- Software Tool – Python 3.6 or above.
- Coding Language – Python.
- Anaconda (Python Distribution)
- Python: Anaconda typically includes the latest version of Python (Python 3.x).
- Jupyter Notebook: Web-based interactive computing environment.
- Operating System Compatibility: Windows, macOS, Linux.
- Jupyter Notebook (Interactive Computing Environment)
- Visual Studio Code (Integrated Development Environment)

Chapter 4

METHODOLOGY

4.1 System Analysis:

The system analysis for developing an automated evaluation system for handwritten answer scripts involves a comprehensive approach to understanding requirements, assessing existing processes, and defining technical specifications. Initially, requirements are gathered through stakeholder interviews and workshops to identify specific needs and objectives, documenting both functional and non-functional requirements essential for system development. A thorough analysis of the current manual evaluation process is conducted to pinpoint inefficiencies and challenges, such as time constraints and subjectivity in grading. A feasibility study is then undertaken to assess technical, economic, and operational feasibility of implementing the automated system, considering factors like technology availability and impact on existing workflows. Design considerations focus on defining system scope, architecture, and scalability requirements to accommodate varying volumes of exam papers during peak periods.

Data collection requirements are specified, including types of handwritten scripts and sources for training data, with preprocessing techniques determined for image cleaning and enhancement. Algorithm selection and development involve evaluating machine learning models suitable for text recognition and evaluation tasks, establishing criteria for assessing answer scripts based on content relevance, correctness, and depth of explanation. User interface design incorporates stakeholder feedback to create intuitive interfaces for paper submission, assessment viewing, and result access, ensuring usability and accessibility for teachers and administrators. Security and privacy considerations are addressed by identifying risks associated with handling sensitive exam data and implementing encryption protocols and access controls.

The strategies are defined to ensure system functionalities, accuracy, and performance are validated under various scenarios through unit testing, integration testing, and user acceptance testing. Comprehensive documentation is prepared, including system requirements, design specifications, implementation guidelines, and user manuals to support system administrators and evaluators. Findings from the system analysis phase are summarized in a formal report, outlining key insights, recommendations, and next steps for successful system development and deployment.

4.2 System Architecture:

The architectural design gives the description about the overall system design. It is specified by identifying the components defining the control and data flow between them. The arrow indicates the connection and rectangular box represents the functional units. The Fig. 4.2.1 shows the architectural diagram of this project which shows the overall operation from uploading question papers to the evaluation of the student answer scripts.

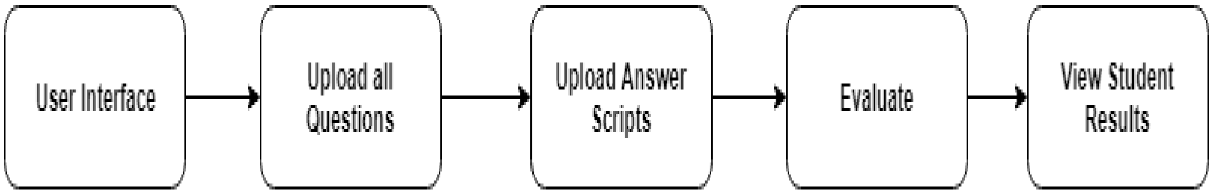


Fig. 4.2.1: Architectural Design Diagram

4.3 Data Flow diagram:

A data flow diagram is a graphical representation of the "flow" of data through an information system, modelling its process aspects. It is often used as preliminary step to create an overview of the system, which can later be elaborated.

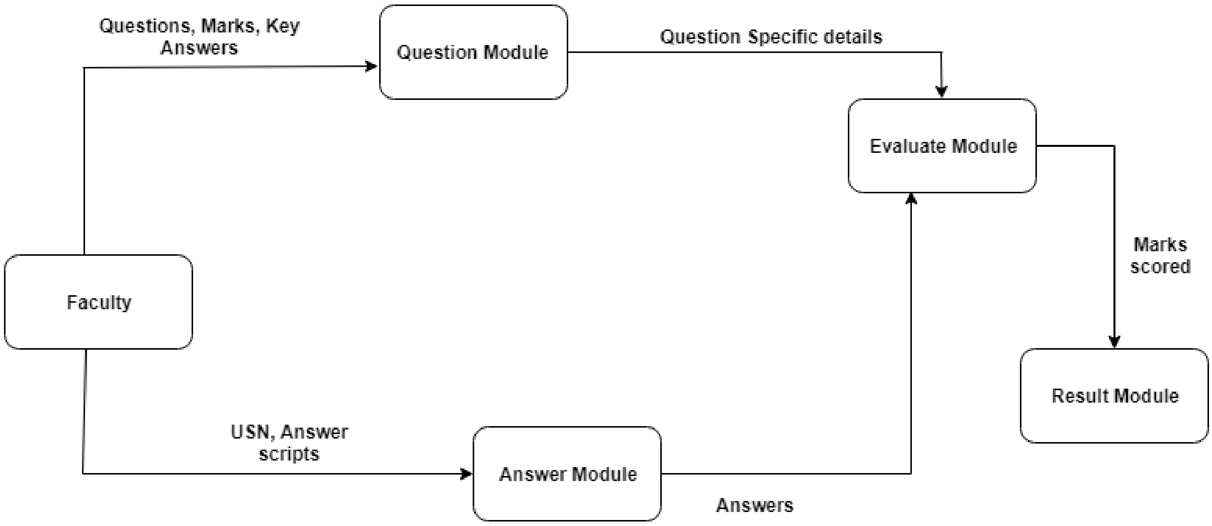


Fig. 4.3.1Data Flow Diagram

The Fig. 4.3.1 shows the data flow between each component in the system. Faculty uploads the question papers and key answers to the Question module. Faculty also uploads the answer scripts to the answer module. These processed answer scripts are given to the evaluate module. Then the results are displayed in the result module.

4.4 Use Case diagram and description:

A use case is a coherent piece of functionality that a system provides interacting with actors. It describes a system which involves a set of use cases and a set of actors. The Fig.4.4.1 shows the use case of the uploading question paper phase where the faculty has to enter the question related details.

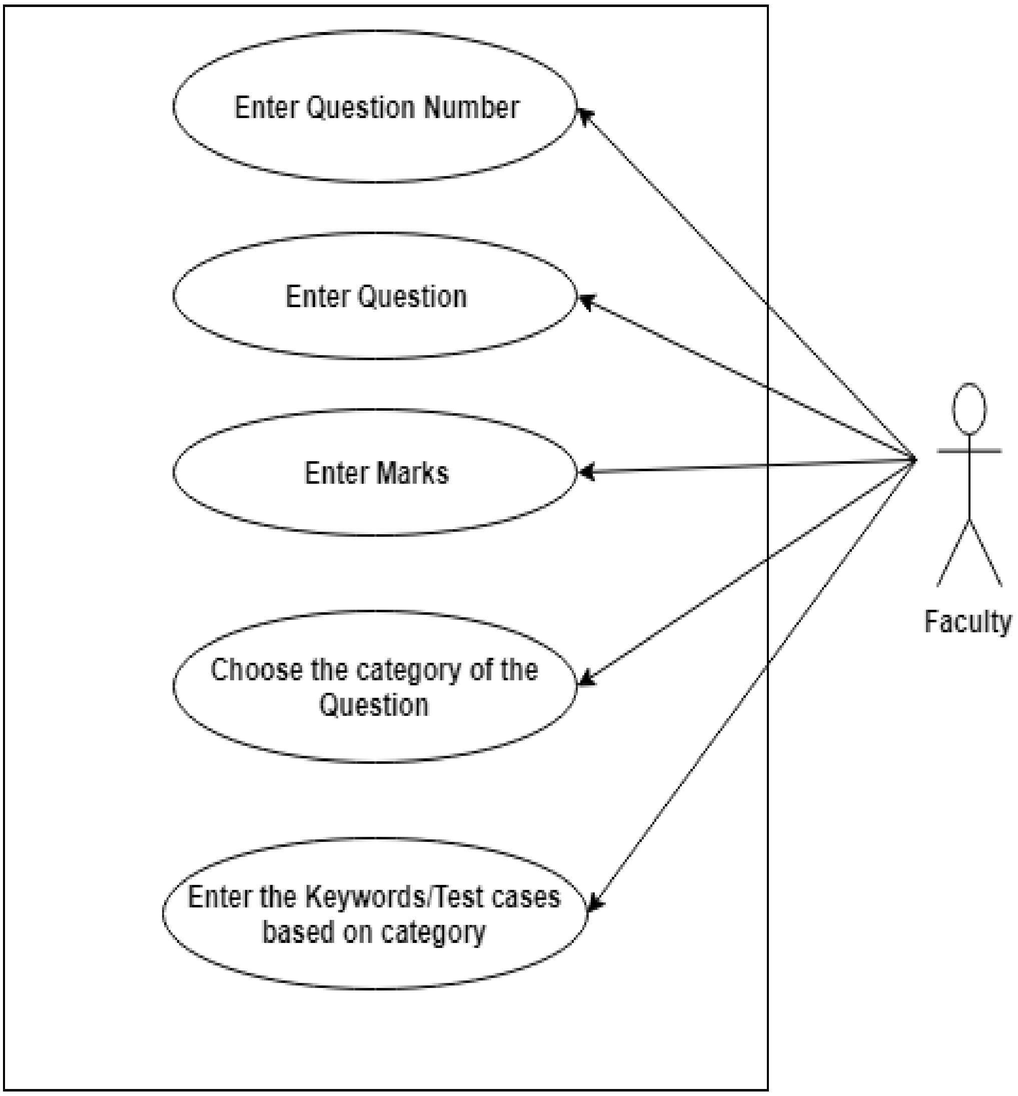


Fig. 4.4.1: Use case diagram for Question paper

The Fig.4.4.2 shows the use case of the evaluation process where the student answer scripts are evaluated based on the criteria provided and displays the result.

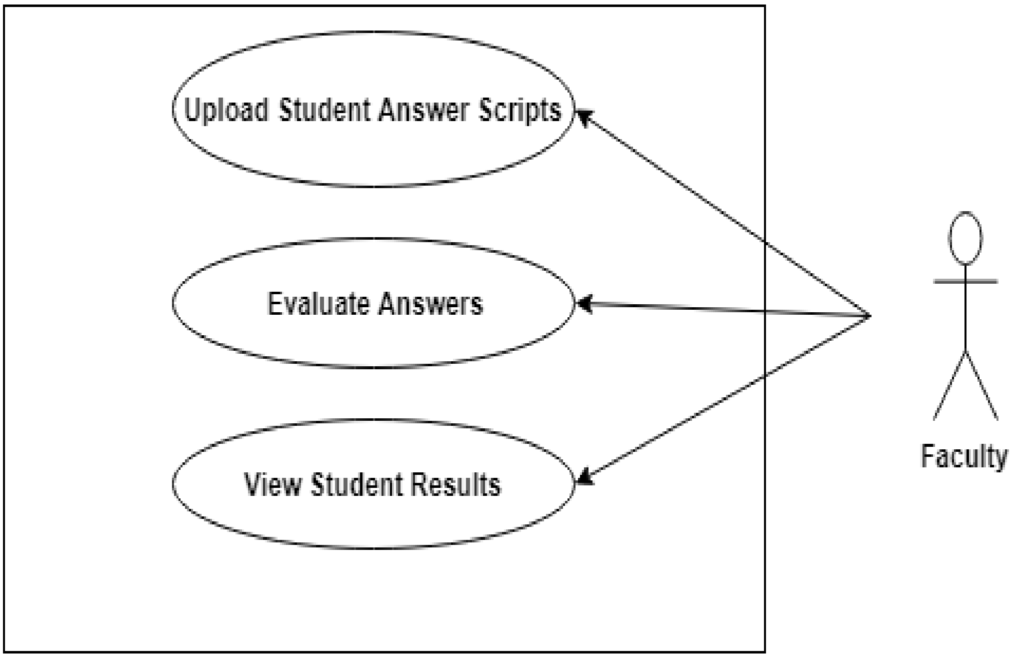


Fig. 4.4.2: Use case diagram for Evaluation

4.5 Activity Diagram:

Activity Diagram shows the sequence of steps that make up complex process. It shows the flow of control, similar to sequence but focuses on operations rather than on objects. The components used in this are as follows:

- Rounded Rectangle indicates the process
- Arrow indicates the transition line
- Rhombus indicates the decision.
- Bars represents the start or end concurrent activities
- Solid circle represents the initial state of work flow
- Encircled black circle represents the final state of work flow.

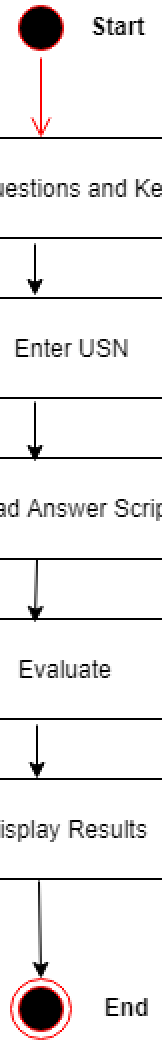


Fig.4.5.1 Activity Diagram

An activity diagram serves as a visual representation of the sequence of steps within a complex process, highlighting the flow of control and operations rather than focusing on specific objects. Key components within this diagram include rounded rectangles, which denote individual activities or tasks within the process. Arrows indicate the transition between different activities, illustrating the logical progression from one step to the next. Decision points are represented by rhombus shapes, where conditions or criteria determine the path the process will take. Bars signify the start or end of concurrent activities, showcasing points in the process where multiple actions may occur simultaneously. The solid circle denotes the initial state of the workflow, marking the starting point before any actions have taken place. Together, these components provide a clear and concise depiction of the entire process, aiding in understanding and analysis of complex workflows.

4.6 Sequence diagram:

A sequence diagram shows how a set of objects communicate with each other to complete a complex task. The Fig. 4.6.1 shows the sequence of operations between the different modules involved in the project.

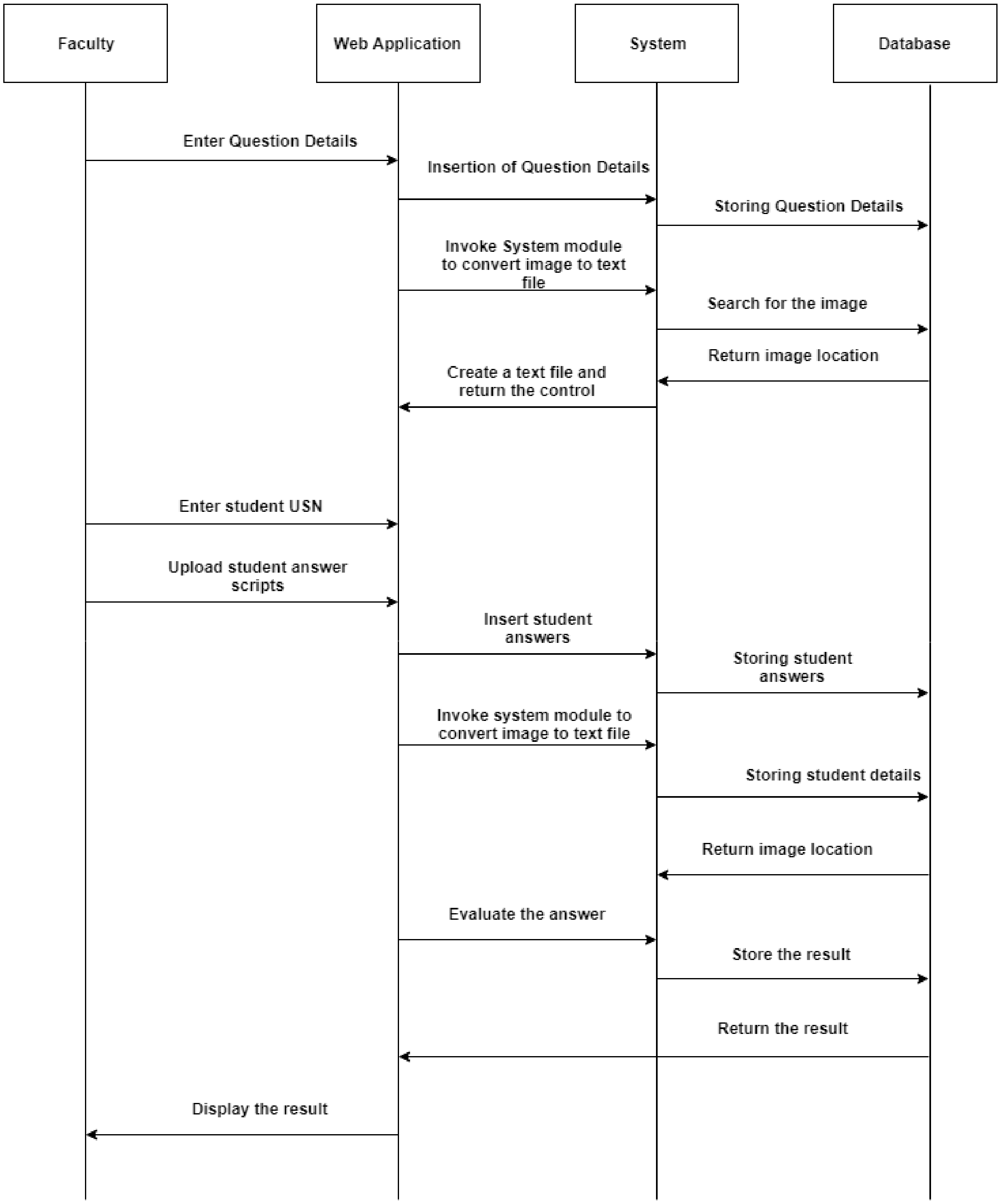


Fig. 4.6.1: Sequence Diagram

Chapter 5

IMPLEMENTATION

Implementation is the core step in software development life cycle. Implementation gives the detailed view of the project and describes the pseudo code and various important functions in the project.

5.1 Algorithm:

An algorithm is a step-by-step instruction to execute the program. The following Algorithm 5.1.1 shows the high-level algorithm of the project.

Step 1: Start

Step 2: Upload a question to the system

1. Input the question number

2. Enter the question

3. Enter the total marks

4. Choose the question category

a. Theory

- *Enter the important keywords*
- *Upload the reference answer in image format*

b. C Program

- *Enter 3 test cases*

Step3: Update question details into the database and convert reference answer from image to text

Step 4: Repeat Step 2 to 3 for all the questions

Step 5: Upload student answer script in image format

Step 6: Convert student answer from image to text

Step7: Evaluate student answer and Update the marks

Step8: Stop

Algorithm 5.1.1: Algorithm of the project

5.2 Code Snippets

5.2.1 Draw the boxes for detected text

```
img = cv2.imread(IMAGE_PATH)
# Image size with DPI
plt.figure(figsize=(20, 20), dpi=100)
i = 0
for detection in result:
  # Top left coordinate
    top_left = tuple([int (val) for val in detection[0][0]])
  # Bottom right coordinate
    bottom_right = tuple([int (val) for val in detection[0][2]])
  # Text extraction
    text = detection[1]
    print(str(i)+' '+text)
  # Draw Rectangle
    img = cv2.rectangle(img,top_left,bottom_right,(0,0,0),2)
    i+=1

plt.imshow(img)
plt.savefig('Detected_Text.png')
plt.show()
```

The following code snippet is used to draw the box around the multiple text lines.

5.2.2 Prediction score Histogram

```
data = []
for i in range(len(result)):
  # Append the prediction score in the list
  data.append([result[i][2]])
  # Create the dataframe of the prediction score
  df = pd.DataFrame(data,columns=["Prediction_Score"])

display(df.head())
df.plot.hist(bins=20)
plt.savefig("Prediction_Score.png", dpi=100)
```

The following code snippet is used to predict the prediction score histogram

.

5.2.3 Crop and store the detection bounds

The following code snippet is used to crop and store the detected bounds

```
for i in range(len(result)):
  X= int(result[i][0][0][0])                # //Column
  Y= int(result[i][0][0][1])                # //Row
  W= int(result[i][0][1][0])                # //Width
  H= int(result[i][0][2][1])                # //Height
# Slicing of particular boxes
  cropped_image = img[Y:Y+H, X:X+W]
# Save an image
  cv2.imwrite(str(i)+'.png', cropped_image)
```

```
# Zip the patches Folder
# /Destination /Source
!zip -r /content/Patches.zip /content/Patches
```

5.2.4 Removing stop words

The following code is used to remove the stop words

```
# REMOVING THE STOPWORDS

stop = stopwords.words('english')

data["anchor"] = data['anchor'].apply(lambda x: ' '.join([word for word in x.split() if word not in (stop)]))
data["target"] = data['target'].apply(lambda y: ' '.join([word for word in y.split() if word not in (stop)]))
```

5.2.5 Preprocess answer scripts

The following code is used to preprocess the student answer.

```
# REMOVING WHITESPACE, HASHTAGS. HTML TAGS AND PUNCTUATION

preproc = preprocessing.make_pipeline(
    preprocessing.remove.punctuation,
    preprocessing.normalize.whitespace,
    preprocessing.replace.hashtags,
    preprocessing.remove.html_tags
 )

data["anchor"] = data["anchor"].apply(preproc)
data["target"] = data["target"].apply(preproc)
```

5.2.6 Cosine similarity

The following code is used to calculate the cosine similarity

```
from scipy import spatial

cos_sim_vals  = []

for a, t in zip(anchor_tf, target_tf):
    similarity = 1 - spatial.distance.cosine(a, t)
    cos_sim_vals.append(similarity)
```

```
data["Cosine Similarity"] = cos_sim_vals
```

5.2.7 Importing Libraries

```
# IMPORTING NECESSARY LIBRARIES

from textacy import preprocessing
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
import nltk
from nltk.stem import WordNetLemmatizer
from nltk.stem.snowball import SnowballStemmer
from gensim.models import Word2Vec
import gensim
```

Chapter 6

TESTING

Testing is an activity to check whether the actual results match the expected results. Testing also helps to identify errors, gaps or missing requirements in contrary to the actual requirements. Testing is an important phase in the development life cycle of the product. During the testing, the program to be tested was executed with a set of test cases and the output of the program for the test cases was evaluated to determine whether the program is performing as expected. Errors were found and corrected by using the following testing steps and correction was recorded for future references. Thus, a series of testing was performed on the system before it was ready for implementation. An important point is that software testing should be distinguished from the separate discipline of Software Quality Assurance (SQA), which encompasses all business process areas, not just testing

6.1 Design of test cases:

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

According to the ANSI/IEEE 1059 standard, Testing can be defined as “A process of analyzing a software item to detect the differences between existing and required conditions and to evaluate the features of the software items”. A good testing program is a tool for agency and integrator supplier it identifies the end of the “Development” phase of the project, establishes the criteria for project acceptance, and establishes the start of the warranty period.

6.2 Testing levels:

Testing is part of Verification and Validation. Testing plays a very critical role for quality assurance and for ensuring the reliability of the software.

The objective of testing can be stated in the following ways.

- A successful test is one that uncovers as-yet-undiscovered bugs.

- A better test case has high probability of finding un-noticed bugs.
- A pessimistic approach of running the software with the intent of finding errors. Testing can be performed in various levels like unit test, integration test and system test.

6.3 Types of testing:

There are many types of testing that can be carried out. Few of the testing types are:

- Unit Testing
- Integration Testing
- Functional Testing
- System Testing
- Acceptance Testing

Unit Testing: Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application, it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results. Unit testing tests the individual components to ensure that they operate correctly. Each component is tested independently, without other system component. This system was tested with the set of proper test data for each module and the results were checked with the expected output. Unit testing focuses on verification effort on the smallest unit of the software design module.

- A testing is carried out to check whether the database is connected to the user interface.
- Testing was done on question uploading unit to check whether the question specific details like question number, question, maximum marks, question category, key answers and image of the reference answer are stored in the local storage.
- A test was carried out on handwriting recognition module to verify whether the images are getting converted into text.

- Testing was carried out on different units of this system like checking the grammar mistakes of student answers, finding the length of the answer, evaluating the similarity between student answer and reference answer, searching for important keywords in student answer.

Integration Testing: Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components. Integration testing is another aspect of testing that is generally done in order to uncover errors associated with the flow of data across interfaces. The unit-tested modules are grouped together and tested in small segment, which makes it easier to isolate and correct errors. This approach is continued until we have integrated all modules to form the system as a whole.

- After the completion of each module it has been combined with the remaining module to ensure that the project is working properly as expected.
- An integration test was performed on evaluation module whether the student answer scripts are uploaded and evaluated, which consist of different units like image to text unit, a unit to preprocess the student answer, a program to check the grammar and find the length of the student answer, a unit to estimate the similarity between student answer and reference answer to check.

Functional Testing: Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

- Valid Input: identified classes of valid input must be accepted.
- Invalid Input: identified classes of invalid input must be rejected.
- Functions: identified functions must be exercised.
- Output: identified classes of application outputs must be exercised.
- Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests are determined.

System Testing: System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration-oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points. System testing tests a completely integrated system to verify that it meets its requirements. After the completion of all the module they are combined together to test whether the entire project is working properly.

- It deals with testing the whole project for its intended purpose. In other words, the whole system is tested here.
- System testing was carried out by uploading questions' details through user interface to the system and then student answer scripts were uploaded and evaluated. Then the system generated students' result which is accessed by the faculty through the user interface.

Acceptance Testing: Project is tested at different levels to ensure that it is working properly and was meeting the requirements which are specified in the requirement analysis.

- Acceptance testing is done once the project is done and checked for the acceptance by uploading student answer scripts to the system for automatic evaluation
- The results from the system was compared with the results from the traditional evaluation approach.
- Then the system was tested in terms of accuracy and throughput against traditional approach.

White Box Testing:

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing:

Black Box Testing is testing the software without any knowledge of the innerworkings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box, you cannot "see" into it. The test provides inputs and responds to outputs without considering how the software works.

6.4 Test Cases

A test case is a vital component in the realm of software testing, serving as a document that meticulously outlines specific conditions and sequences of actions to verify a particular aspect of the software application. Each test case typically includes several elements such as events, actions, inputs, expected outputs, and actual outputs. These elements collectively help in assessing whether the software behaves as intended under various scenarios. A well-structured test case provides a detailed test description, clear procedure steps, anticipated results, and space for remarks or comments. This structured approach ensures that every functional aspect of the software is thoroughly validated against its requirements.

Primarily, test cases are derived from the software requirements specifications (SRS), which document the intended functions and features of the application. By aligning test cases with these requirements, testers ensure that the software meets its designed specifications and performs its intended functions correctly. This alignment also helps in identifying and establishing conditions that might reveal potential errors, inconsistencies, or gaps in functionality. A comprehensive set of test cases will cover various scenarios, including edge cases and unusual input conditions, to robustly test the software's behavior and resilience.

For each test case, individual PASS/FAIL criteria are explicitly defined. These criteria serve as the benchmarks for determining the success or failure of the test. If the actual output matches the expected output, the test case is marked as PASS, indicating that the software functions correctly for that specific scenario. Conversely, a mismatch between the actual and expected results leads to a FAIL status, signaling a defect or issue that needs to be addressed. This binary evaluation system simplifies the tracking of testing progress and the identification of problematic areas within the application.

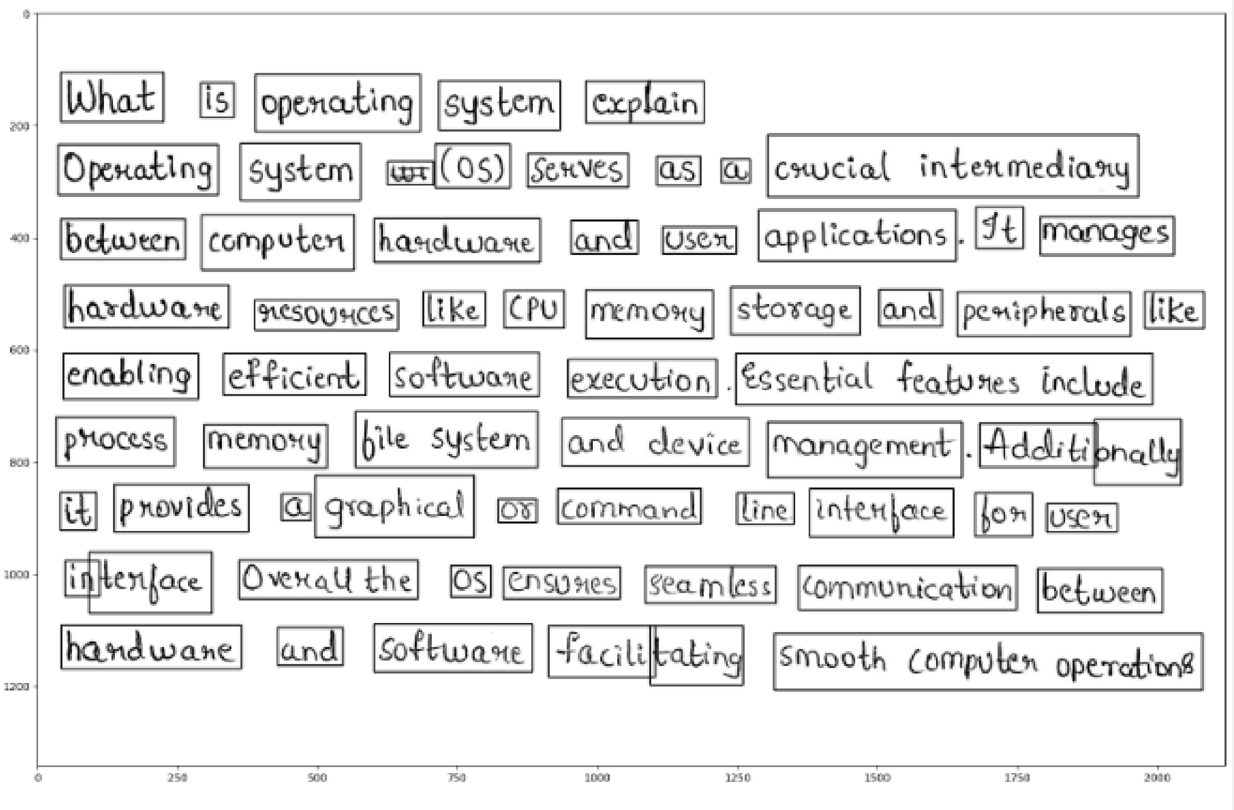
Ensuring that all test cases result in a PASS is crucial for the proper functioning of the application. It indicates that the software has been rigorously tested and meets its quality standards and requirements. This comprehensive testing process is essential not only for detecting defects but also for validating that the software delivers a reliable and satisfactory user experience. Therefore, meticulous development and execution of test cases form the backbone of effective software testing, enabling the delivery of robust, error-free applications to end users.

The test Cases are summarized as shown in Table 6.4.1

Test Numbers	**Test Case ID**	**Test Case**	**Expected Results**	**Status**
1	UT_1	Uploading inputs into system	Question specific details are stored in the spread sheet	Uploaded
2	UT_2	Uploading answer script into system	Student answer has to be uploaded to the system	Uploaded
3	UT_3	Converting image to text	The image has to be converted into a text file	Converted
4	UT_4	Finding similarity between student answer and key answer	The similarity between answer and key answer is calculated out of 100	Evaluated
5	UT_5	Display the match result	To display the match results	Displayed
6	UT_6	Clear answer script from the spread sheet	To delete previous data	Deleted
7	UT-7	Delete questions from the system	To delete the questions from the system	Deleted

Table 6.4.1

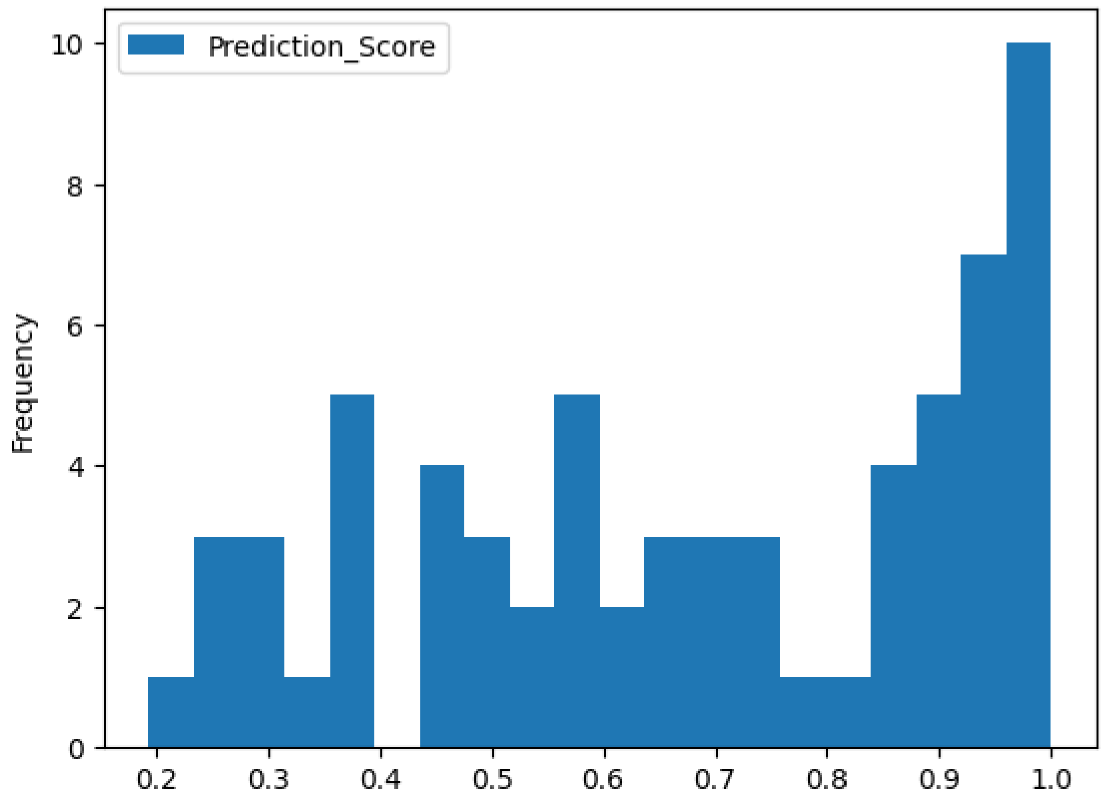
RESULTS AND SNAPSHOTS



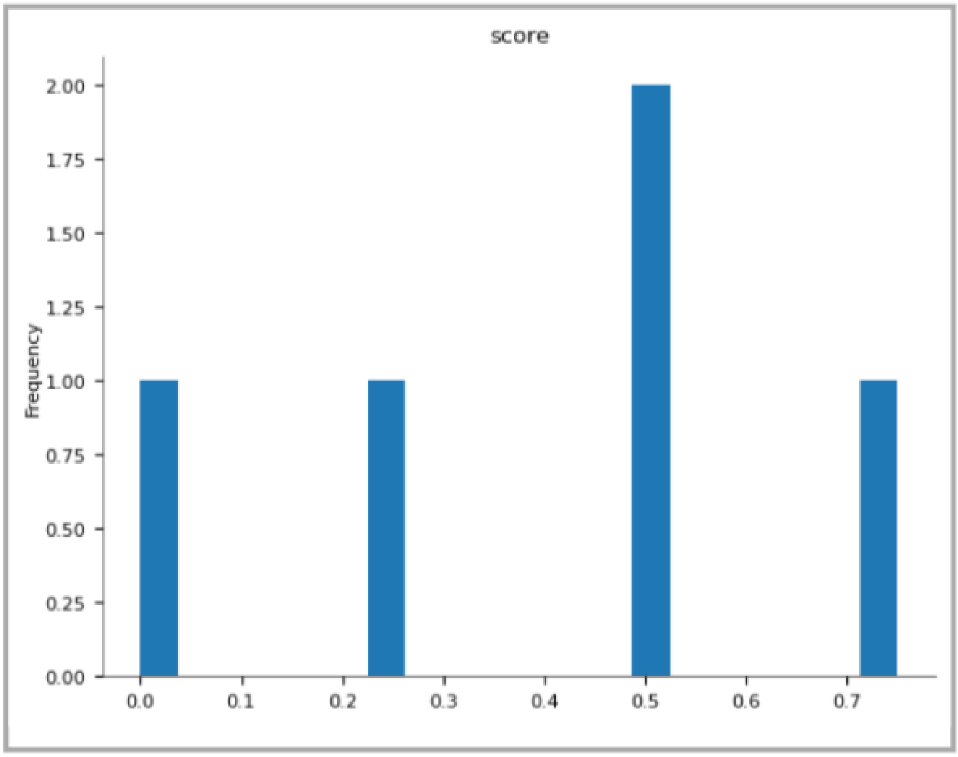
Snapshot 7.1 The text extraction using OCR

	Prediction_Score
count	66.000000
mean	0.678553
std	0.247326
min	0.194230
25%	0.476054
50%	0.687677
75%	0.920223
max	0.999898

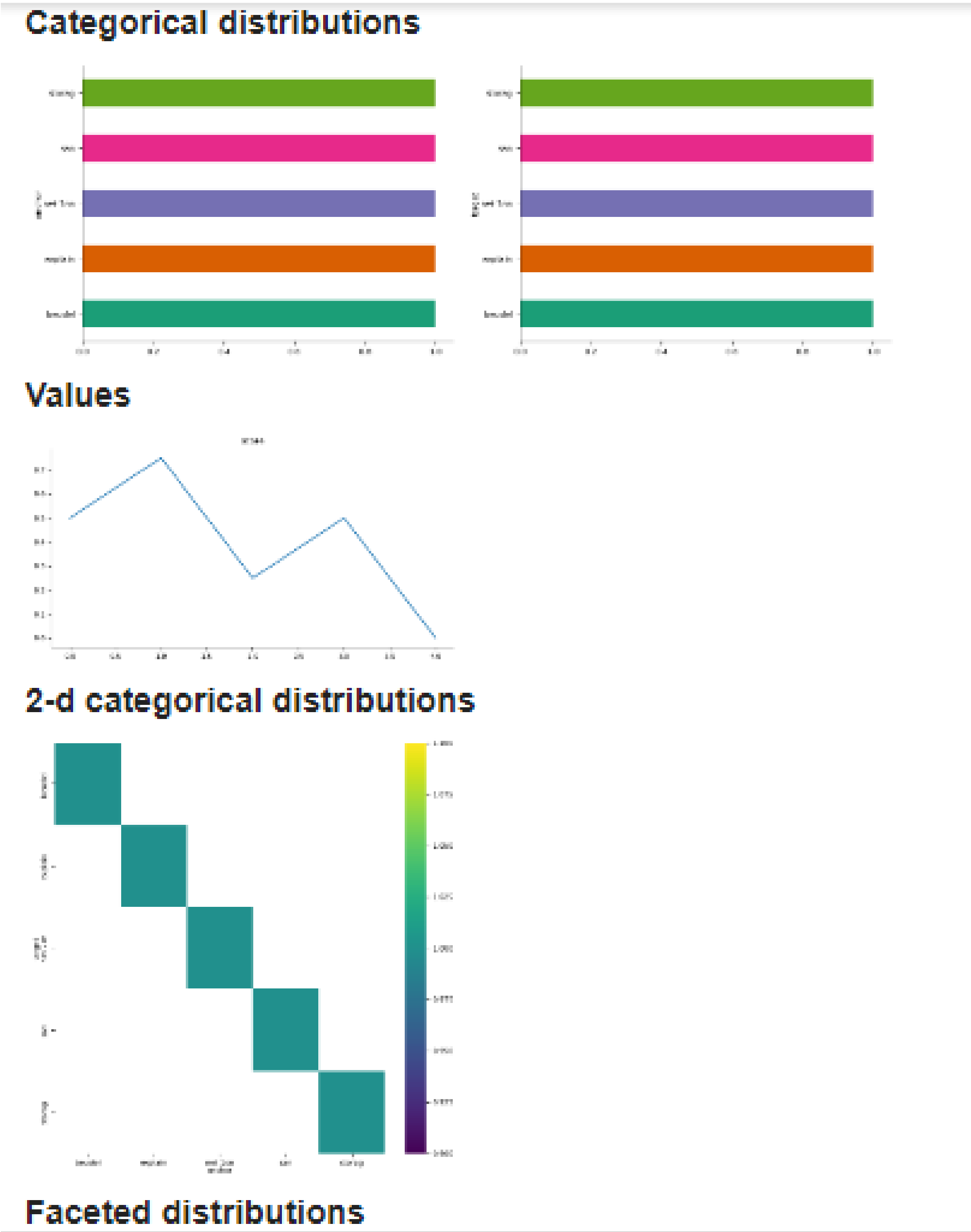
Snapshot 7.2 Prediction Score



Snapshot 7.3 Prediction Graph



Snapshot 7.4 Score Distribution



Snapshot 7.5 Data Visualization, Analytical Overview, and Graphical Summary

CONCLUSION

In conclusion, the development and implementation of an automated evaluation system for handwritten answer scripts mark a significant step forward in the field of education assessment. By harnessing the power of machine learning algorithms and computer vision techniques, this project has successfully addressed the challenges associated with manual evaluation processes, such as time consumption and subjective biases. The automated system offers a more efficient and objective means of grading answer scripts, thereby streamlining the assessment process and providing timely feedback to students. The automated evaluation system holds immense potential for enhancing the overall quality and effectiveness of education.

Leveraging advanced technologies to assess handwritten answer scripts, educators can devote more time to teaching and mentoring students, rather than spending hours on manual grading tasks. This shift towards automation not only improves the efficiency of educational institutions but also fosters a more conducive learning environment, where students receive prompt and constructive feedback on their academic performance. The continued refinement and deployment of automated evaluation systems have the potential to revolutionize education assessment practices on a global scale.

As technology continues to evolve, future iterations of the automated system may incorporate additional features and functionalities, such as natural language processing algorithms for assessing essay-type questions or adaptive learning techniques for personalized feedback. Ultimately, the integration of automated evaluation systems into educational institutions represents a significant step towards achieving greater efficiency, fairness, and transparency in the assessment process, ultimately benefiting students, educators, and educational institutions alike.

FUTURE ENHANCEMENT

In advancing the automated evaluation system for handwritten answer scripts, future enhancements can focus on combining advanced handwriting recognition algorithms with sophisticated natural language processing techniques. This integration would enable the system to assess more complex responses, such as essays and open-ended questions, by analyzing both the content and coherence of handwritten text. By expanding the dataset used for training and incorporating machine learning models trained on diverse handwriting styles and languages, the system can enhance its accuracy in interpreting various handwriting patterns, ensuring robust performance across different demographics and contexts.

Furthermore, leveraging cloud-based infrastructure and distributed computing technologies can enhance the scalability and accessibility of the automated evaluation system. Deploying the system on cloud platforms would enable educational institutions to access its capabilities on-demand, without the need for extensive on-premises infrastructure. This scalability ensures that the system can efficiently handle large volumes of answer scripts during peak assessment periods, while also accommodating future growth and expansion. Additionally, incorporating feedback mechanisms and adaptive learning algorithms into the system can personalize the assessment experience for individual students, providing tailored feedback and recommendations to address each student's unique learning needs and fostering a more supportive and engaging learning environment.

REFERENCES

1. Engelhard, G., Jr. (2002). Monitoring raters in performance assessments. Large–scale assessment programs for all examinees: Validity, technical adequacy, and implementation (pp. 261–287). Mahwah, NJ: Lawrence Erlbaum.
2. Wang, Z., & Yao, L. (2013). Investigation of the effects of scoring designs and rater severity on students' ability estimation using different rater models (Research Report. No. RR–13–23). Princeton, NJ: Educational Testing Service.
3. Zhang, M. (2013, March). Contrasting automated and human scoring of essays. (R & D Connections, No. 21). Princeton, NJ: Educational Testing Service.
4. K. Zechner, D. Higgins, X. Xi, and D. M. Williamson, "Automatic scoring of non-native spontaneous speech in tests of spoken English," Speech Communication, vol. 51, pp. 883-895, 2009.
5. Bennett, R. E., & Bejar, I. I. (1998). Validity and automated scoring: It's not only the scoring. Educational Measurement: Issues and Practice, 17(4), 9–17.
6. Attali, Y., Bridgeman, B., & Trapani, C. (2010). Performance of a generic approach in automated scoring. Journal of Technology, Learning, and Assessment, 10(3). Retrieved from www.jtla.org
7. Chodorow, M., & Burstein, J. (2004). Beyond essay length: Evaluating e-rater®'s performance on essays (Research Report No. RRí04í04). Princeton, NJ: Educational Testing Service.
8. Wang, Z., & von Davier, A. A. (2014). Monitoring of scoring using the e-rater automated scoring system and human raters on a writing test (Research Report No. RR–14í04). Princeton, NJ: Educational Testing Service.
9. Bachman, L. F. (1990). Fundamental considerations in language testing. Oxford: Oxford University Press.
10. F. Dong and Y. Zhang, "Automatic features for essay scoring – an empirical study," In Proc. of EMNLP, pp. 1072–1077, 2016.
11. K. Taghipour and H. T. Ng, "A neural approach to automated essay scoring," In Proc. of EMNLP, pp.1882–1891, 2016.

12. D. Alikaniotis, H. Yannakoudakis, and M. Rei, “Automatic text scoring using neural networks,” in Proc. of ACL, vol. 1, pp. 715– 725, 2016.
13. B. Riordan, A. Horbach, A. Cahill, T. Zesch and C.M. Lee, “Investigating neural architectures for short answer scoring” in Proc. of the 12th Workshop on Innovative Use of NLP for Building Educational Applications, pp. 159-168, 2017.
14. C. M. Lee, S.-Y. Yoon, X. Wang, M Mulholland, I. Choi and K. Evanini, “Off-topic spoken response detection using Siamese Convolutional Neural Networks,” in Proc. of Interspeech, 2017.
15. A. Malinin, K. Knill, A. Ragni, Y. Wang and M. Gales., “An attention based model for off-topic spontaneous spoken response detection: an Initial Study,” in Proc. of SLaTE, 2017.
16. Z. Yu, V. Ramanarayanan, D. Suendermann-Oeft, X. Wang, K. Zechner, L. Chen, J. Tao, A. Ivanou, and Y. Qian, in Proc. of ASRU, pp. 338–345, 2015.
17. Dong, F., Zhang, Y. (2016). Automatic features for essay scoring–an empirical study. In Proceedings of the2016 Conference on Empirical Methods in Natural Language Processing (pp. 1072-1077).
18. Shehab, M. Elhoseny and A. E. Hassanien, ”A hybrid scheme for Automated Essay Grading based on LVQ and NLP techniques,” 2016 12th International Computer Engineering Conference (ICENCO), 2016
19. Balci, D. Saadati, and D. Shiferaw, Handwritten Text Recognition using Deep Learning, Stanford University Ray Smith, Google Inc., “An Overview of the Tesseract OCR Engine”
20. Chirag Patel, Atul Patel and Dharmendra Patel, “Optical Character Recognition by Open Source OCR Tool Tesseract: A Case Study”, International Journal of Computer Applications (0975 – 8887) Volume 55– No.10, October 2012
21. Shivani Dhiman and A.J Singh, “Tesseract Vs Gocr A Comparative Study”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-2, Issue-4, September 2013.
22. G. R. Perera, D. N. Perera and A. R. Weerasinghe, “A dynamic semantic space modelling approach for short essay grading,” 2015 Fifteenth International Conference on Advances in ICT for Emerging Regions (ICTer), Colombo
23. M. Jangid and S. Srivastava, “Automatic Objective Answer-Sheet Evaluation-A OCR Based Approach1,” 2013 IEEE International Conference in MOOC,Innovation and Technology in Education (MITE), Jaipur, 2013, pp. 224-227

24. T. Gunawardena, M. Lokuhetti, N. Pathirana, R. Ragel and S. Deegalla, “An automatic answering system with template matching for natural language
25. https://en.wikipedia.org/wiki/Natural_language_processing
26. https://pypi.org/project/pytesseract/
27. https://www.paperrater.com
28. https://www.pyimagesearch.com/2017/07/10/using-teseract-ocr-python
29. https://en.wikipedia.org/wiki/Software_requirements_specification

International Research Journal Of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

e-ISSN: 2582-5208

Ref: IRJMETS/Certificate/Volume 06/Issue 05/60500233176

Date: 25/05/2024

Certificate of Publication

*This is to certify that author "**Nithin BK**" with paper ID "**IRJMETS60500233176**" has published a paper entitled "AUTOMATED EVALUATION OF HANDWRITTEN ANSWER SCRIPT USING DEEP LEARNING APPROACH" **in International Research Journal Of Modernization In Engineering Technology And Science (IRJMETS), Volume 06, Issue 05, May 2024***

IRJMETS Impact Factor 7.868

Editor in Chief

We Wish For Your Better Future

www.irjmets.com

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:05/May-2024 Impact Factor- 7.868 www.irjmets.com

AUTOMATED EVALUATION OF HANDWRITTEN ANSWER SCRIPT USING DEEP LEARNING APPROACH

Siddaraj MG*1, Sukesh P*2, Nithin BK*3, Shreyas MS*4, Abhishek V Hadapad*5

*1Assistant Professor, Department Of Information Science And Engineering, Maharaja Institute Of Technology Mysore, Mysore, Karnataka, India.

*2,3,4,5Undergraduate Student, Department Of Information Science And Engineering, Maharaja Institute Of Technology Mysore, Mysore, Karnataka, India.

ABSTRACT

This machine learning project focuses on automating handwritten answer script evaluation using advanced processing techniques. Employing Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), it extracts meaningful features for consistent feedback across subjects. Challenges like handwriting recognition, grading, and feedback generation are addressed, promising to revolutionize assessments by providing timely, objective evaluations. It reduces manual grading burdens, enhancing educational efficiency. Its adaptability to diverse subjects and educational levels highlights its versatility. Future enhancements may include real-time monitoring, advancing automation. This innovation addresses manual grading challenges, ensuring efficient, objective assessment in education.

Keywords: Automatic Script Grading, Bidirectional LSTM Network, Convolutional Neural Network, Deep Learning, Natural Language Processing.

I. INTRODUCTION

Since the last four decades, computer is using for writing essay and to assess automatic submissions of students work homogeneously. The best pedagogy for improving students writing skills is to check each submission for each student and to replay them individually by a teacher in the classroom. Unfortunately, this significantly increases the workload for teachers. In fact, responding to student papers and rigorously checking them is a burden for many teachers and this pressure linearly increases with the number of students increases. Therefore, developing an automated system can help to reduce the cost of checking in a significant way and facilitates students to get early feedback. Handwritten text recognition is the ability of a computer to interpret a text 1 from sources like scripts, images, or others. The image is scanned optically with handling formats, segmenting lines and words into characters to trace the most plausible characters. The most challenging problem related to handwriting recognition is recognizing different styles and sizes with a good accuracy level. The aim of this study is to explore the task of classifying handwritten text and to convert them into digital format and to grade them automatically. The development of AEG was mostly started with Latent Semantic Analysis (LSA), N-Gram, TF-IDF, Bayesian classifier, and K-nearest neighbor approaches, although the satisfactory performance level wasn't achieved well. After the revaluation of Deep Learning (DL) and Natural Language Processing (NLP), much research has been done on the automatic evaluation of computer-based submit ted essays and higher accuracy is gathered.

II. LITERATURE REVIEW

[1] Engelhard, G., Jr. (2002). Monitoring raters in performance assessments. Large–scale assessment programs for all examinees: Validity, technical adequacy, and implementation (pp. 261–287). Mahwah, NJ: Lawrence Erlbaum.

In educational settings, as performance assessments diversify, monitoring and evaluating rating quality becomes imperative, especially in systems incorporating constructed-response items. Rater-mediated (RM) assessments involve raters judging examinee responses using a rating scale, where responses serve as stimuli for raters to interpret and evaluate. It's crucial that measurement models used for RM assessments accurately reflect rater behavior, performance, and response. RM assessments, relying on raters' interpretations, don't offer direct insights into examinee achievement. Concerns arise due to potential biases among raters, impacting the validity of test score interpretations, particularly in state assessment and accountability systems. Maintaining fairness and validity amidst potential biases is essential for accurate assessment.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:05/May-2024 Impact Factor- 7.868 www.irjmets.com

[2] Wang, Z., & Yao, L. (2013). Investigation of the effects of scoring designs and rater severity on students' ability estimation using different rater models (Research Report. No. RR–13–23). Princeton, NJ: Educational Testing Service.

The paper from IFET College of Engineering presents an OCR system using Recurrent Neural Networks (RNN) for handwritten English text recognition. It addresses challenges in diverse handwriting styles, employing RNN layers and Connectionist Temporal Classification (CTC) for accuracy. Divided into four stages—Dataset collection, Preprocessing, Training and Testing Set division, and RNN implementation—it achieves 90% accuracy in recognizing characters and digits. Python and TensorFlow are used for implementation. The paper outlines each stage meticulously, showcasing a comprehensive approach. Achieving 90% accuracy signifies the system's efficacy, with room for further improvement. This reflects dedication to innovation in OCR technology, promising more accurate solutions.

[3] Zhang, M. (2013, March). Contrasting automated and human scoring of essays. (R & D Connections, No. 21). Princeton, NJ: Educational Testing Service.

Essay scoring traditionally relies on human raters who understand content and writing quality. But with the rise of constructed-response items, especially in assessments like the Common Core State Standards (CCSS), there are concerns about relying solely on human scoring. Human scoring is costly, logistically challenging, and subjective. Hence, testing programs are exploring computerized scoring for efficiency. However, human and automated scoring have distinct strengths and limitations. Research, including ETS studies, compares these methods from measurement and logistical standpoints. Despite debates in academia and the media, there's limited comprehensive research comparing both methods' advantages and limitations. Test developers, policymakers, and educators need a nuanced understanding of each method's strengths and weaknesses to avoid misuse. This essay aims to contrast significant characteristics of human and automated scoring, highlight differences, and discuss practical implications for testing programs.

[4] K. Zechner, D. Higgins, X. Xi, and D. M. Williamson, "Automatic scoring of non-native spontaneous speech in tests of spoken English," Speech Communication, vol. 51, pp. 883- 895, 2009.

This paper introduces SpeechRaterSM, the first system for automatically scoring non-native spontaneous high-entropy speech, designed for the Test of English as a Foreign Language® internet-based test (TOEFL® iBT). It comprises a speech recognizer trained on non-native English data, a feature computation module for fluency-based features, and a regression scoring model predicting speaking proficiency. Experiments with classification and regression trees complement multiple regression. Evaluation on TOEFL Practice data and Field Study data shows a correlation of 0.57 with human scores, warranting deployment in low-stakes practice environments. The system achieves 57.8% exact agreement with human scores on TOEFL Practice Online Speaking tests since 2006, indicating its operational viability. Future iterations will incorporate vocabulary, grammar, and content features as speech recognition technology advances.

[5] Bennett, R. E., & Bejar, I. I. (1998). Validity and automated scoring: It's not only the scoring. Educational Measurement: Issues and Practice, 17(4), 9–17.

Early automated scoring work relied on translating conventionally delivered tasks into machine-readable form due to limited computer-based testing options. This emphasized empirical characteristics of automated scores. With the rise of computer-based testing, operational exams could be fully implemented, broadening discussions on automated scoring's validity implications. This paper asserts that validity discussions should encompass construct definition, test and task design, examinee interface, tutorial, test development tools, automated scoring, and reporting, as these components interact during development. Modern validity theory suggests empirical evidence of score relationships alongside theoretical rationales supporting design decisions. The interdependency among computer-based test components presents an opportunity for significant improvement in educational and occupational assessment.

III. METHODOLOGY

The architectural design serves as a blueprint for the entire system, detailing its components and the flow of control and data between them. Each component is identified and represented, illustrating how they interact. Arrows signify connections, while rectangular boxes denote functional units. The architectural diagram of the project outlines the entire process, from uploading question papers to evaluating student answer scripts. This

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:05/May-2024 Impact Factor- 7.868 www.irjmets.com

diagram provides a visual representation of the system's operation, allowing stakeholders to understand the workflow at a glance. It helps in clarifying the system's structure, enabling developers to identify potential bottlenecks or areas for optimization. Moreover, it serves as a communication tool, facilitating discussions among team members and stakeholders regarding system functionality and requirements. Overall, the architectural design plays a crucial role in guiding the development and implementation of the project, ensuring its success and effectiveness in achieving its objectives.

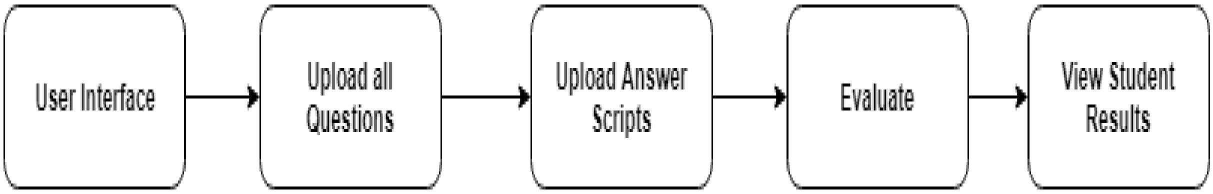


Figure 1: Architecural design of proposed system

A. Data Flow Diagram

Figure 2 illustrates the intricate data flow among the various components within the system. Initially, faculty members upload question papers and corresponding key answers to the Question module, initiating the assessment process. Simultaneously, they upload answer scripts to the Answer module, which serves as the repository for student responses. Subsequently, these answer scripts undergo processing within the Evaluate module, where they are systematically evaluated based on predefined criteria or key answers provided earlier. This evaluation stage is critical as it determines the accuracy and fairness of grading. Finally, the evaluated results are conveyed to the Result module, where they are formatted and displayed for easy access by relevant stakeholders, such as faculty and students. This seamless flow of data ensures the efficient operation of the system, enabling timely and accurate assessment of student performance while providing valuable insights for both educators and learners.

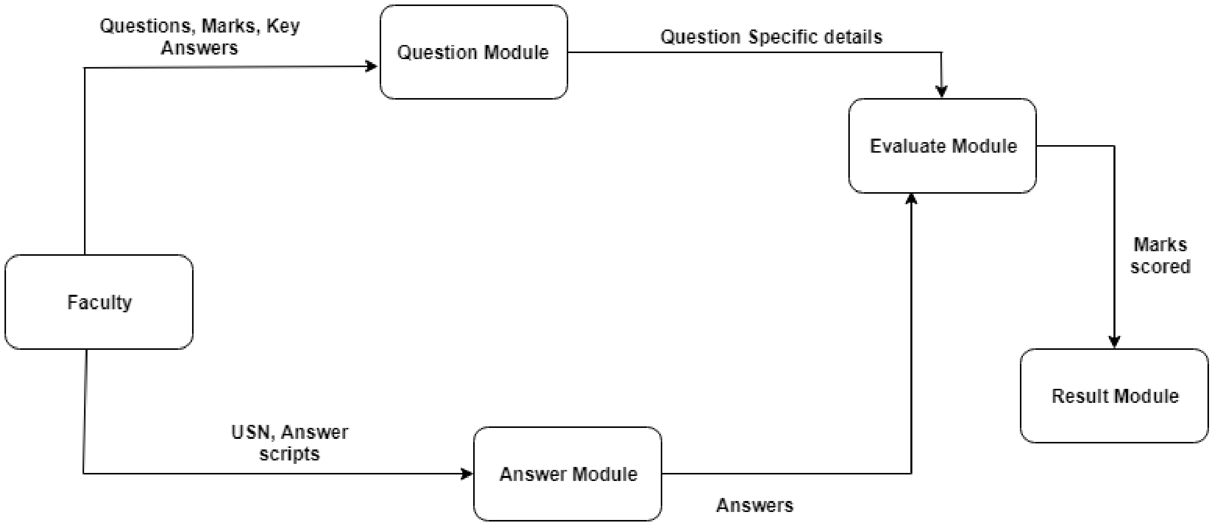


Figure 2: Data Flow Diagram

IV. EXPERIMENTAL RESULT

The result of the automated evaluation system for subjective answers involves comparing the student's response with a predefined key answer provided by the university. Marks are allocated based on how closely the student's answer matches the model answer. This approach saves time and effort for educational institutions, eliminating the need to manually check numerous papers. The system grades answers based on 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.

V. CONCLUSION

The development and implementation of an automated evaluation system for handwritten answer scripts marks a significant advancement in education assessment. By utilizing machine learning algorithms and computer vision techniques, this project effectively addresses the challenges of manual evaluation, including time constraints and subjective biases. The automated system streamlines grading processes, providing efficient and objective assessment while offering timely feedback to students. It holds great potential for enhancing education quality by allowing educators to focus on teaching rather than grading. Automation fosters a conducive learning environment where students receive prompt feedback. As technology evolves, future iterations may integrate additional features like natural language processing for essay-type questions or adaptive learning for personalized feedback. Ultimately, automated evaluation systems enhance efficiency, fairness, and transparency in assessments, benefiting students, educators, and institutions globally

VI. FUTURE SCOPE

Future enhancements of the automated evaluation system for handwritten answer scripts can involve integrating advanced handwriting recognition algorithms with sophisticated natural language processing techniques. This integration would enable the system to analyze complex responses like essays and open-ended questions, assessing both content and coherence. Expanding the training dataset and incorporating diverse handwriting styles and languages would improve accuracy across demographics. Leveraging cloud-based infrastructure and distributed computing technologies would enhance scalability and accessibility, allowing institutions to access the system on demand. Feedback mechanisms and adaptive learning algorithms could personalize the assessment experience, providing tailored feedback and recommendations to address individual learning needs, thereby creating a more supportive and engaging learning environment.

VII. REFERENCES

[1] Engelhard, G., Jr. (2002). Monitoring raters in performance assessments. Large–scale assessment programs for all examinees: Validity, technical adequacy, and implementation (pp. 261–287). Mahwah, NJ: Lawrence Erlbaum.

[2] Wang, Z., & Yao, L. (2013). Investigation of the effects of scoring designs and rater severity on students' ability estimation using different rater models (Research Report. No. RR–13–23). Princeton, NJ: Educational Testing Service.

[3] Zhang, M. (2013, March). Contrasting automated and human scoring of essays. (R & D Connections, No. 21). Princeton, NJ: Educational Testing Service.

[4] K. Zechner, D. Higgins, X. Xi, and D. M. Williamson, "Automatic scoring of non-native spontaneous speech in tests of spoken English," Speech Communication, vol. 51, pp. 883- 895, 2009.

[5] Bennett, R. E., & Bejar, I. I. (1998). Validity and automated scoring: It's not only the scoring. Educational Measurement: Issues and Practice, 17(4), 9–17.

[6] Attali, Y., Bridgeman, B., & Trapani, C. (2010). Performance of a generic approach in automated scoring. Journal of Technology, Learning, and Assessment, 10(3). Retrieved from www.jtla.org

[7] Chodorow, M., & Burstein, J. (2004). Beyond essay length: Evaluating e-rater®'s performance on essays (Research Report No. RRí04í04). Princeton, NJ: Educational Testing Service.

[8] Wang, Z., & von Davier, A. A. (2014). Monitoring of scoring using the e-rater automated scoring system and human raters on a writing test (Research Report No. RR–14í04). Princeton, NJ: Educational Testing Service.