IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 1, January 2025

DOI: 10.17148/IJARCCE.2025.14108

A Survey: ML-Based Automated Handwriting Analysis and Answer Evaluation

Sarthak Karmalkar[1], Ansari Siraj[2], Shakti Singh[3], Mrudula kulkarni[4],

Prof. Naved Raza Q.Ali[5], Prof. Dhanashri Nevase[6]

Undergraduate Survey Paper, Department of Computer Engineering, SKNCOE, Savitribai Phule Pune University, Pune 411041, India[1,2,3,4]

Department of Computer Engineering, SKNCOE, Savitribai Phule Pune University, Pune 411041 India[5,6]

Abstract: In a technologically advancing world, the evaluation of answers should happen rapidly and with greater accuracy. However, unlike objective answers, subjective answers make it difficult for an automated system to evaluate them accurately. This is because subjective answers are hard to evaluate using static content and finding a dynamic capability that caters to content, meaning, order and structure for subjective type answer evaluation is not so easy. This study represents an automated evaluation system for handwritten as well as textual answer sheets making use of ML and NLP for the evaluation. This survey is all about a system that converts the answers written on the answer sheets into their digital text data, then check whether answer of each question is correct or not. This study comprises of various "Machine Learning" algorithm to recognize and digitize text from handwritten forms. It also analyzes the answer of a student based on keyword matching, semantic similarity and correct grammar and according to that it assigns marks for their given answer using various "Machine Learning" techniques and algorithms. These systems help to minimize biased marking scheme and promotes fair grading. Also, ensuring consistent evaluation and less human work. An overview has been provided, which includes its evolution and effectiveness of various Machine Learning (ML) techniques to improve "Subjective answer evaluation systems".

Keywords: Optical Character Recognition (OCR), Convolutional Neural Networks (CNN), Machine learning (ML), Natural Language Processing (NLP), Large Language Models (LLM), Subjective Answer Assessment.

I. INTRODUCTION

Evaluating handwritten subjective answers has always been an essential part of any organization, whether it's a college, school, company, or other institution. Examinations can be either descriptive or objective or both. Every examination needs evaluation. The majority of competitive exams are objective in structure. Hence, they are easy to evaluate. These methods cannot be used in board exams or university exams where students give subjective answers due to a few issues, hence there is a need for software that will automate evaluation for subjective answers which includes handwritten or textual data.

In the current digital era, handwriting recognition has become more and more important, as there is an increasing demand for recognizing handwritten characters and digits in many areas. In short, Optical Character Recognition (OCR) [1] systems are indispensable when it comes to automating the data processing workflow stages and using the man-hours otherwise spent on manual typing for other tasks. Though OCR has traditionally been used to convert typewritten documents into a machine-readable format, recent advances have moved the field of recognition in handwriting use. Handwriting recognition has the task of handling arbitrary writing styles, which may include many specialized symbols in addition to total variation in character shapes and sizes. These complexities drive the need for more complex machine learning models and larger data sets to achieve better recognition. Significant advances have been observed in offline handwriting recognition due to the emergence of machine learning techniques such as "Convolutional Neural Networks (CNN)" [2].

The key contribution of the technologies mentioned in these papers is a framework for translating hand-written answers to their text equivalent, and then checking if it deviates from the model answer in terms of both its accuracy and relevance. Various technologies are surveyed for subjective answer evaluation which includes 'Cosine similarity', 'Jaccard similarity', 'Fuzzy Wuzzy', etc. [3]. These papers elaborate the utilization of the custom dataset for fine tuning which

DOI: 10.17148/IJARCCE.2025.14108

includes many answers to questions and model answer for that corresponding question, along with the evaluation and grading of that answer.

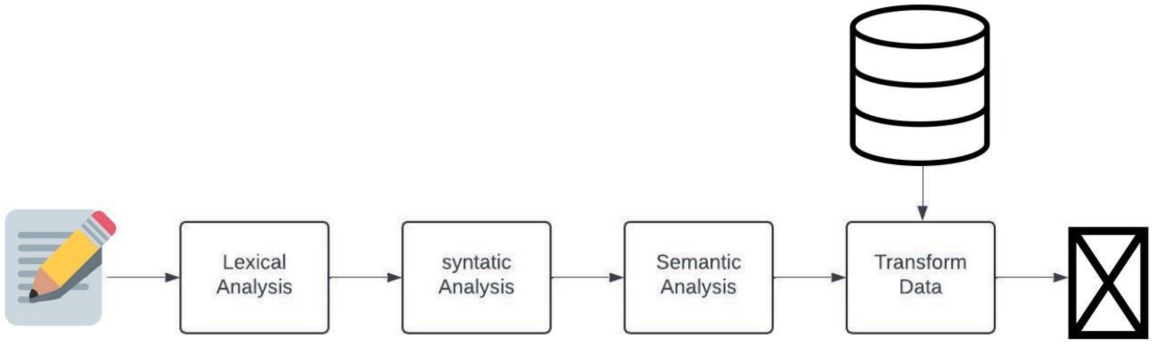

Fig. 1: System architecture of handwriting recognition

It is a broader system that evaluates subjective responses like text similarity, keyword. We apply NLP techniques such as tokenization to convert both the student's answer and the model answer into a similar format. These representations can then be merged using “LLMs” like “GPT3”, “BERT” [13]. Automatic subjective evaluation systems intendeds minimal human intervention and the system lets assessments be far more objective, scalable and efficient. This functionality promotes a time-saving, resource-saving, unbiased academic assessment tool especially in the digital era where most of the education frameworks are transforming [14].
These papers present a system that utilizes the ‘NLP and BERT’ for handwriting recognition, along with ‘Machine Learning’ to evaluate the content [15].

Section 1 of the survey paper contains Abstract, section 2 of the paper contains Introduction, section 3 contains Literature Survey, section 4 contains Observation, section 5 contains Conclusion, section 6 contains Future Scope and Section 7 contains References.

II. LITERATURE SURVEY

This literature survey explores recent advancements in AI- driven automated answer evaluation systems, focusing on methodologies like OCR, neural networks, and large language models.
Preetha S. et al. [1] presented handwriting recognition, with a special focus on CNN for handwritten character recognition. She and her team discuss about seven techniques for handwriting recognition. Namely, CNN, Incremental method, Semi incremental method, Slope and slant, Line and word segment, Part based method, Ensemble method. She also mentioned about OCRs as they lack accuracy.

Jamshed Menon et al. [2] presented comprehensive review on OCRs for handwriting recognition between 20092019. The paper summarized various machine algorithms used for Text recognition like, support vectors, decision tree, K-nearest neighbour. Among all machine learning algorithm, SVM is most accurate (92-98.4%). This was tested on six most widely spoken language. Further, the paper challenges research in OCR. Also, recognition in real-world settings and requires large amount of dataset size and variety.

Varun Aggarwal et al. [3] presented, the computer-assisted system to automate the assessment of student’s answers to subjective questions. This paper argues that due to traditional assessment methods, a lot of time can go into waste and some selection bias or human bias does exist.
Therefore, authors recommended using similarity algorithms like “cosine similarity”, “Fuzzy Wuzzy”, or “Jaccard similarity” for comparing the answers provided by students to the reference answers. The study was carried out on the data of 100 IT students' answers concerning the APIs. The results indicate that the cosine similarity performed best, with an accuracy of 74.7% compared to that of a human grader. The paper concludes on the importance of having multiple reference answers by the testing center as one way of boosting the accuracy of the automated assessment system.

Sharad Bharadiya et al. [4] proposed the use of Machine Learning for automatic evaluation of answers. The system uses optical character recognition on handwritten answers to compare them with a database of keywords and length parameters

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 1, January 2025

DOI: 10.17148/IJARCCE.2025.14108

given by teachers. This will scale down the timescale and human effort invested in grading subjective questions. Hence, results in faster evaluation and consistent than by hand grading.

Era Johri et al. [5] presented an architecture of subjective answer evaluation using Semantic Learning, Sentence Encoding, Similarity Matrix. Similarity Matrix used by them allowed the generation of feedback for the students answer but at the same time system was not capable of evaluating contradictory answers compared to the model answer.

Farrukh Bashir et al. [6] combined the various methods like NLP, Tokenization, Stop words removal, POS tagging, Lemmatization, Stemming, Case folding, Bag of Words, TF-IDF, Word2Vec, Cosine Similarity, Jaccard Similarity, WMD in order to improve the answer evaluation methods, which rose the efficiency of the system to 88%.

Kavita Shirsat et al. [7] proposed a model to evaluate subjective answers using NLP, Machine Learning, Similarity Index (Cosine Similarity) and Universal Sentence Encoding. The proposed model used 3 factors for evaluating the answers, namely similarity index, grammar and question specific parameters, but small size of dataset used to train the model inhibits the efficiency of the system. Universal Sentence Encoder(USE) converts sentences into 512 dimensional embedding vector, which preserves the actual meaning and context of the answers of the students. Question Specific Parameter searches for a particular keyword in the written paragraphs of the students. Assigned values 0 and 1, based upon the grammar of the answer(0 for bad grammar and 1 for good grammar)

Shreya Singh et al. [8] built a system for the evaluation of handwritten sheets of students by using OCR to convert handwritten answers into digital text, summarizing the answer to further evaluate the responses. The use of Recurrent Neural Network (RNN) improved the image recognition performance, along with algorithms like Word2Vec and TF-IDF to find the importance of every word in the answer.

Nandita Bharambe et al. [9] presented a system that combined both handwriting to digital text conversion and the evaluation of the same. They used supervised learning algorithms like Artificial Neural Network (ANN) and OCR to detect the text present in the handwriting. In ANN, the use of Back-propagation allows the neurons to improve their performance by sending errors to the previous neurons. To evaluate the answers, they used Cosine Similarity, matching the keywords of the model answer and also giving importance to the length of an answer. Marks are assigned out of 10; if the length of an answer is less than the required length, marks are given between 1 to 4. Conversely, if the length exceeds the threshold set by them, then the marks of the student will lie between 5 and 10.

Prerana M S et al. [10] introduce a framework that integrates models like BERT, GPT-3, CNN, LSTM, and SVM with Optical Character Recognition (OCR). Their goals include improving scoring accuracy and scalability, while also addressing challenges such as keyword dependency and the significant resources needed for model training.

Vijay Kumari et al. [11], leverage BERT and TF-IDF to highlight the balance between automation and consistency. Their system basically included two important modules Checker and Evaluator. Their method relies heavily on predefined correct answers, which can create difficulties when faced with ambiguous responses.

Sheik Abdullah et al. [12], concentrate on OCR, NLP, and Connectionist Recurrent Neural Networks (CRNN), addressing issues related to handwriting variability and the challenges of data collection. Their research offers valuable insights into the complexities involved in automating the grading process. In [13], they mentioned OCR, NLP, and ML are used to automate answer script grading. It helps increase accuracy and efficiency but has issues in terms of handwriting variability and the inability to evaluate the non-text components properly. Open- ended questions need more development for their proper evaluation as well as diverse formats of answering scripts Further, Vaibhav Shikhar Singh et al. [14] share their findings where the system achieves an accuracy rate of 83.14% by utilizing artificial neural networks (ANN) for OCR tasks, although they recognize ongoing challenges related to handwriting quality. Collectively, these studies demonstrate notable advancements in automatic answer evaluation systems while underscoring the necessity for further research to improve their effectiveness in educational contexts.

Madhavi Kulkarni et al. [15] built a system to evaluate handwritten answer sheets using OCR, NLP, ML, BERT and cosine similarity. OCR can parse handwritten content into a structured format, but not process further before giving the NLP model, which actually performs textual data analysis, helping to match student answers with model answers pre-defined in machine-readable text. So, this provides predictions of new submissions based on analysis of the data. It also

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ≍ Peer-reviewed & Refereed journal ≍ Vol. 14, Issue 1, January 2025

DOI: 10.17148/IJARCCE.2025.14108

creates instant feedback for students highlighting where they may have strayed, in order for the assessment process to be more efficient and precise.

Md. Afzalur Rahaman et al. [16] Implemented a system for evaluation of handwritten sheets of students by using Bidirectional LSTM Network (BiLSTM), Convolutional Neural Networks (CNN). They includes powerful tools such as: Natural language processing (NLP), support vector regression (SVR) and bayesian linear ridge regression (BLRR) to Automatically grade handwritten answer scripts. The system will save educators time, eliminate human error, and be for practice learning and then improve. But even though the system is highly practical, its limitations such as computational complexity along with contextual understanding and recognition challenges have raised several important aspects that need to be addressed. However, today the model gets only around 80% accuracy — far from enough that confidence can be instilled to deploy the system into a production environment for handling complex textual and image-based content.[15]

Sangeeta Mangesh et al. [17] presented Subjective Answer Script Evaluation using Natural Language Processing, in which various techniques were combined, namely NLP, Gaussian Naive Bayes Approach, Machine learning, and Cosine Similarity. The proposed system will generate meaning from textual content using the techniques of NLP. It will classify and predict scores based on NLP features of the text and train it using previously evaluated answers. The system achieves around 80% accuracy and improves by adding some mathematical tools that will assess specific mathematical and chemical equations with precision.

III. OBSERVATIONS AND FINDINGS

In this survey, the authors studied various papers, analyzing key trends, advancements, and challenges within the field. Over the course of these four years, distinct patterns have emerged, reflecting the evolution of methodologies and the increasing focus on specific areas of study. The following section represents the recurring themes, technological developments, and existing research gaps based on the authors' analysis of the reviewed papers. In [5], a mathematical formula was used to calculate the grades of students by encoding their answers using Google's Universal Sentence Encoder, which was not highly accurate. This accuracy was further enhanced by the authors in [6], leading to an improvement to 88% using methods like cosine similarity, Word2Vec, WMD, and Jaccard Similarity, along with Machine Learning to train the model to achieve this accuracy. In
[7] and [8], OCR was utilized to convert students' handwriting into digital text, which was further evaluated using techniques like cosine similarity and keyword matching.
In [10], modern Large Language Models such as GPT-3 and BERT were used to fine-tune the model and evaluate the answers, improving the efficiency of the evaluation process. In [13], Natural Language Processing was employed to extract text from handwritten answer sheets. This analysis indicated that using LLM models and fine-tuning them, along with deep learning algorithms like CNN and ANN, can enhance system performance. In [15], cutting-edge technologies like OCR and NLP were utilized to preprocess and convert handwriting into text, and BERT (Bidirectional Encoder Representations from Transformers) and cosine similarity were used to evaluate students' answers. The same paper presented a mathematical formula: $F = O / (M * Q)$, where 'Q' represents the total number of questions, 'M' signifies the maximum possible score that any student might obtain, and 'O' is the overall score calculated by the weighted sum of 'H' and 'A' (where H = handwriting recognition score and A = content analysis score). This paper thus employed both mathematical and machine learning models.

IV. CORE TAKEAWAYS AND CHALLENGES

One significant advantage is increased accessibility for visually impaired individuals, along with real-time character recognition and versatility in recognizing characters. Techniques like template matching, artificial neural networks (ANN), and recurrent neural networks (RNN) exhibit higher accuracy compared to traditional methods such as SVM, random forests, and decision trees. The use of Fuzzy Wuzzy allows for more flexibility, accounting for minor spelling mistakes.

Efficiency and scalability emerge as major advantages, as these models improve the speed and cost-effectiveness of evaluation systems. Additionally, the implementation of a similarity matrix enables the generation of feedback mechanisms by identifying missing points. Various methods, including the use of similarity indexes, grammar checks, and question-specific parameters, have also improved the performance of answer evaluation models compared to existing systems.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 14, Issue 1, January 2025

DOI: 10.17148/IJARCCE.2025.14108

Another key insight is that the use of RNNs enhances system efficiency for text extraction from answer sheets, while back-propagation and activation functions like Rectified Linear Unit (RELU) help improve neural network performance by reducing feature matrix dimensions. Handwriting recognition (OCR), when integrated with other techniques, contributes to increased accuracy, objectivity, and scalability. The use of multiple techniques, including RAKE and YAKE for rapid keyword extraction, ensures consistency and efficiency in evaluation, while deep learning models like CRNN provide a comprehensive pipeline for OCR and NLP tasks, improving overall system performance.

The combination of automation, deep learning, and semantic understanding enhances the accuracy and reduces the time required for complex tasks. This also leads to better feedback mechanisms, faster responses, and scalable solutions that can integrate large language models (LLMs). Furthermore, features like facial detection and detailed analysis enhance objectivity in specific evaluation scenarios. Overall, these methods provide cost-effective, scalable, and time-efficient solutions for educators, minimizing human error and promoting continuous learning and improvement in evaluation systems.

V. CONCLUSION

This study examines how artificial intelligence and machine learning can be employed to automate the assessment of subjective responses on handwritten documents. The system leverages sophisticated optical character recognition (OCR) methods, including Convolutional Neural Networks and Large Language Models, to digitize and assess handwritten text. The evaluation process is based on matching keywords, analyzing semantic similarities, and checking grammatical correctness.

This approach minimizes manual labor and time investment and introduces elements of fairness, neutrality, and uniformity to the evaluation process. The method offers a scalable and efficient solution compared to conventional techniques, delivering swifter and more unbiased assessments. Initial tests and surveys indicate the system achieves an 80% accuracy rate. As technological advancements continue, this system could be adapted for discipline-specific grading, support for multiple languages, and implementation across various educational domains.

In summary, this project evaluation marks a significant advancement toward creating more dynamic, efficient, and universally accessible academic assessments, paving the way for future educational innovations. existing solutions—such as limited personalization, inadequate sustainability support, and lack of student-specific features— Campus Core aims to improve accessibility, personalization, and functionality tailored to student needs.

VI. FUTURE WORK

The existing system can be enhanced by integrating advanced AI and machine learning models to improve handwriting recognition, especially for complex content like mathematical equations and scientific diagrams. Expanding multi-language support can increase its versatility for diverse educational settings worldwide. Implementing adaptive learning mechanisms can provide personalized, real-time feedback based on individual student performance and learning styles, boosting engagement and outcomes.

Improvements in scalability can ensure the system efficiently handles millions of users, supporting global educational deployments without compromising performance. Strengthening data security and ensuring compliance with international standards will be crucial for protecting sensitive user information. Furthermore, integrating models like GPT-4 and other state-of-the-art natural language processing algorithms will enhance interactivity and engagement, significantly impacting modern education.

REFERENCES

[1] Preetha S, Afrid I, Karthik Hebbar, Nishchay S, “Machine Learning for Handwriting Recognition,” *International Journal of Computer (IJC)* Volume 38, No 1, pp 93-101, 2020.

[2] JAMSHED MEMON, MAIRA SAMI, RIZWAN AHMED KHAN, “Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR),” *IEEE ACCESS* r 10.1109/ACCESS.2020.3012542 VOLUME 8, Aug. 2020.

[3] Varun Aggarwal, Rutuja Sutar, Shweta Tiwari, Pankaj Choudhary “Automated Assessment of Students Responses to the Questions using Various Similarity Techniques,” *International Journal of Engineering Research & Technology (IJERT)* ISSN: 2278-0181 Published by, www.ijert.org ICSITS - Volume 8, Issue 05, Dec. 2020.

[4] Sharad Bharadia, Prince Sinha, Ayush Kaul, Dr. Sheetal Rathi “Answer Evaluation Using Machine Learning,” *Researchgate* https://www.researchgate.net/publication/333856264_Answer_Evaluation_Using_Machine_Learning, 2020.

DOI: 10.17148/IJARCCE.2025.14108

[5] Era Johri, Nidhi Dedhia, Kunal Bohra, Prem Chandak, Hunain Adhikari “ASSESS-Automated Subjective Answer Evaluation using Semantic Learning,” *Social Science Research Network (SSRN), 2021*

[6] Farrukh Bashir, Hamza Arshad, Abdul Rehman Javed, Natalia Kryvinska, Shahab S. Band “Subjective Answer Evaluation
Using Machine Learning and Natural Language Processing,” *Institute of Electrical and Electronics Engineers(IEEE)* DOI:10.1109/ACCESS.2017.DOI, 2021

[7] Kavita Shirsat, Akash Hamare, Sourabh Pardeshi, Vaishnavi Patil “Subjective Answer Evaluation System,” *Internation Research Journal of Engineering and Technology(IRJET),* volume 8, 2021

[8] Shreya Singh, Omkar Manchekar, Ambar Patwardhan, Prof. Uday Rote, Prof. Sheetal Jagtap, Dr. Hariram Chavan “Computer Application for Assessing Subjective Answers using AI,” *Social Science Research Network(SSRN), 2021*

[9] Nandita Bharambe, Pooja Barhate, Prachi Dhannawat “Automatic Answer Evaluation Using Machine Learning,” *International Journal of Information Technology(IJIT)* volume 7, 2021

[10] Prerana M. S., Sagarika M. Chavan, Ramit Bathula, Sreenath Saikumar, and Geetha Dayalan, "Eval - Automatic Evaluation of Answer Scripts using Deep Learning and Natural Language Processing," *International Journal of Innovative Science and Engineering (IJISAE*)

[11] V. Kumari, P. Godbole, and Y. Sharma, "Automatic Subjective Answer Evaluation," *International Conference on Pattern Recognition Applications and Methods (ICPRAM)* 2023

[12] A.Sheik Abdullah, S. Geetha, A.B. Abdul Aziz, Utkarsh Mishra School of Computer Science and Engineering, Vellore Institute of Technology Chennai “A pedagogical approach with NLP and Deep Learning,” *Alexandria Engineering Journal (AEJ),* 2024

[13] S. K. Virdi, H. Wadhwani, D. Hegde, A. Bhardwaj, and S. N,” Automated Answer Sheet Evaluation And Grading System”. *Department of Artificial Intelligence and Machine Learning, (IJNRD)* | Volume 9, Issue 06, June 2024

[14] V. S. Singh, A. Verma, G. Srivastava, and S. Kumar, "An Automated System for Efficient Answer Sheet Evaluation,” *Asian Journal of Research in Computer Science (AJRCOS),* vol. 17, no. 6, April 2024.

[15] Madhavi Kulkarni, Gayatri Adhav, Kunal Wadile, Vrushali Deshmukh, Rohini Chavan “Digital Handwritten Answer Sheet Evaluation System” *International Journal of Computer Applications* (*IJCA*) (0975 – 8887) Volume 186 – No.16, 2024

[16] Md. Afzalur Rahaman, Hasan Mahmud “Automated Evaluation of Handwritten Answer Script Using Deep Learning Approach” *Research Gate*, 2022

[17] Sangeeta Mangesh, Prateek Maheshwari, Aditi Upadhyaya “Subjective Answer Script Evaluation using Natural Language Processing” *JETIR* Volume 9, Issue 8 (JETIR) August 2022

[18] Prof. Sharayu Lokhande, Prof Shilpa Pawar, Udit Chaudhary
,Akash Singh, Pranay Gaikwad ,Himanshu Guleria “Automated Subjective Answer Evaluation System”, Journal Of Algebraic
Statistics Volume 13, No. 3, 2022

[19] Ishika Aggarwal, Pallav Gautam, Gaurav Parashar, “Automated Subjective Answer Evaluation Using Machine Learning” (SSRN), 2024

[20] António J. R. Neves, Daniel Lopes: A practical study about the Google Vision API. Research gate Conference: 22nd Portuguese Conference on Pattern Recognition, RECPAD 2016