IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

Review on Smart Evaluation of Descriptive Answer Sheets

Akash Kiran S[1], Amruta Madev Poojari[2], Dr. Vimuktha E Salis[3]
Department of Information Science and Engineering[1,2,3]
Global Academy of Technology, Bangalore, Karnataka, India
akashkiran0710@gmail.com

Abstract: *This descriptive abstract summarizes a thorough examination into the use of smart technology for answer sheet evaluation. The study explores how to automate the grading process using Artificial Intelligence, Machine Learning and other algorithms to improve efficiency and objectivity while evaluating student responses. Examined are several smart assessment systems, stressing attributes such as adaptive learning processes, pattern recognition and natural language processing. The abstract delves into the possible advantages, obstacles and ramifications linked to the implementation of intelligent response sheet assessment techniques in educational environments. The abstract offers insights into the changing landscape of assessment methodologies through a synthesis of recent research findings, illuminating the revolutionary potential of intelligent systems in reshaping education in the future.*

Keywords: Smart assessment, Natural Language Processing, Grading process, Pattern recognition

I. INTRODUCTION

The context and justification for investigating cutting-edge technologies in the evaluation and grading of student responses are covered in this survey of smart answer sheet assessment. The manual grading of answer sheets in the traditional educational paradigm has long been associated with the risk of subjectivity and time-consuming nature. Education professionals and organizations are looking to creative ways to update and improve the assessment process as they become aware of its shortcomings.

Artificial intelligence (AI) and machine learning (ML) are being used more and more in education as we move into a time of fast technological growth. A manifestation of this evolution is the advent of smart answer sheet evaluation, which aims to use modern technologies to transform the way assessments are administered. Smart evaluation systems seek to address long-standing issues by improving and automating the grading process, potentially leading to evaluations that are more impartial, accurate, and efficient.

The review explores the essential elements and characteristics of these systems within the larger context of smart answer sheet evaluation. Algorithms for machine learning and artificial intelligence are essential for interpreting written responses using natural language processing (NLP), seeing patterns to find the right answers, and customizing the assessment procedure for each student.

This technologically advanced method improves the learning experience for students by speeding up the grading process and creating opportunities for individualized feedback.

Additionally, the exploration on how flexible smart assessment systems is in a variety of learning environments. These systems are flexible tools for teachers since they can be customized to fit different curriculum areas, educational levels, and test styles. By allowing feedback to be given immediately and customized to each student's needs, this technology not only transforms the grading process but also opens the door to a more individualized learning experience for students.

The major highlights are the advantages but also notes the difficulties in integrating smart answer sheet evaluation. The adoption and deployment of these technologies must carefully consider several important factors, including ethical considerations, the requirement for thorough training for instructors, and guaranteeing the security and privacy of student data.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

The background portion also delves into the rationale and benefits of implementing intelligent assessment systems in learning environments. This covers factors like the capacity to manage high assessment numbers effectively, deliver timely and helpful feedback, and promote a more uniform and impartial grading procedure. At the same time, it recognizes the difficulties that come with this revolutionary change, including moral dilemmas, the need for teacher preparation, and the significance of protecting student privacy and security.

To sum up, this new technique to assessment has a thorough overview of its technological foundations, historical context, and driving forces in the introduction and background of smart response sheet evaluation. This basis lays the groundwork for a deeper investigation and examination in the areas of the study that follow.

II. LITERATURE REVIEW

In [1], research aims to develop applications that automatically generate short answer questions in the reading comprehension section of the IELTS test. The computational model is based on previous research and uses systems like DeconStructure, dependency, SRL parse, TextRank, and internal NLU analysis methods. The goal is to reduce error rates and improve overall performance. The process involves collecting articles, preprocessing them into sentences, and converting them into a tree structure using the Stanford Core NLP library. The Elementary Sentence Extracted System (ESES) is used to extract elementary or simple sentences from complex ones. The process is crucial for creating an effective IELTS reading comprehension question.

The [2] paper introduces a system that automatically evaluates handwritten answer booklets using Deep Learning and Natural Language Processing techniques, generating text, summary, and marks based on extracted keywords. Himani et al. developed a technique for extracting terms from sentences for information retrieval and effective searching. The study uses named entity recognition and syntactic grammar techniques to identify glossary terms. Bojja et al. suggest a deep convolutional neural network, encoder-decoder, and attention mechanism for character identification. Sueiras et al. propose a text summarization technique. The paper discusses the performance of handwriting recognition models due to the lack of commercially available data. Cloud-based solutions like Google Cloud Vision and AWS Textract OCR models have shown significant improvements in accuracy and error rates. The paper also investigates keyword extraction and summarization using BERT and GPT3. This paper proposes a system to automate the evaluation process for professors using a combination of Deep Learning and Natural Language Processing. The system uses scanned answer booklets and expected answers to generate suggested marks and summary of the given answer. The system has three stages: image to text conversion, processing the answer, and marks allocation. The extracted text is saved as text files and processed using the NLP model. The extracted keywords are then used in marks allocation to assign marks to the student's answer. The system's overall output includes the answer, summary, and final allocated marks. The system aims to reduce downtime caused by hardware or software failures.

[3] Arabic handwritten-text recognition uses OCR and deep learning techniques, with CNN-LSTM-CTC architecture being most suitable for OCR. DL models improve accuracy, but more high-quality datasets are needed for future Arabic OCR applications.

The OCR process involves several major steps. Image acquisition involves obtaining a photograph from an external source, while pre-processing involves techniques like noise removal, thinning, and binarization. Segmentation involves splitting the image into characters, words, or lines, using techniques like component analysis and projection profiles. Feature extraction helps in learning the representation, while classification predicts the mapping between characters and their classes. Post-processing involves further corrections to reduce errors, using techniques like dictionary creation, probabilistic models, and natural language processing techniques.

This [4] paper proposes a computerized system for Question Generation and Answer Assessment (QGAA), which can grade student written responses and generate subjective questions from course curriculums. The system uses a Multi-criteria Decision Making (MCDM)-based approach to evaluate the correctness of subjective answers. He proposed a system that automates subjective examinations by generating and evaluating answers automatically. It generates subjective questions using Jaccard similarity coefficient, extracts meaningful keywords, and creates templates. Subject experts collect model answers for evaluation. A weighted multi-criteria-decision-making (MCDM) approach is used for final decision-making. The system evaluates responses using similarity measures and preprocessing tasks like grammar checkers and Stanford Parser.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

This [5] paper proposes an automated text analysis-based approach for automatic evaluation of descriptive answers in examinations, utilizing Natural Language Processing and Data Mining concepts. The architecture creates a sample answer sheet, calculates final scores using text summarization, semantics, and keywords summarization, and uses a Siamese Manhattan LSTM (MaLSTM) text similarity model. This paper introduces a computational intelligence-based method for automatic evaluation of descriptive answers in education, calculating scores based on correctness, language, grammar, and size. The proposed system uses a pipeline structure to evaluate student answers. The examiner inputs questions and corresponding evaluation factors based on a survey conducted at an institution. The input includes an ideal answer, the number of works, and required keywords. The system calculates the total score using NLP and text analysis-based intelligent models. Scoring parameters include the size of the answer, language of the answer, presence of necessary keywords, and similarity index of submitted answer and ideal answer. The ideal number of words expected for an answer is considered as x, with a score of 5%. The model also evaluates the structure and grammar of the answer, giving a score of 5%. The system also checks if the student has used the specified keywords in the answer, using word banks and simple if-else methods. The system's focus is on the similarity index between the submitted answer and the ideal answer, using various NLP tools and techniques.

The [6] paper discusses the need for automation in answer assessment systems, highlighting the challenges of manual evaluation and introduces the 'AutoEval' system, which standardizes paper correction and uses natural language processing methods. The paper proposes a digital approach using Natural Language Processing (NLP) for an Automatic Test Evaluation System, aiming to make the answer script evaluation mechanism quicker and more accurate. The system, created during the pandemic, saves time and manpower by automatically scoring points based on similarity between uploaded answers and ideal answers, making it easier to track and extract. The project uses Natural Language Processing (NLP) in Python to analyze descriptive response scripts and assign marks to questions. It uses the Natural Language Toolkit (NLTK) for tokenization, parsing, classification, stemming, tagging, and semantic reasoning. Gensim, a leading NLP package, is used for text document processing and word vector modeling, creating a dictionary object with unique word ids.

This research [7] aims to improve Optical Character Recognition (OCR) performance by addressing challenges like image noise, handwriting styles, font sizes, and alignment. It uses Tesseract, an open-source OCR engine, and Java programming language to analyze digital handwritten images, addressing pixel noise, fuzzy, and grainy data, and utilizing OpenCV for image processing.

This research developed a canvas to create a dataset of digital handwritten text for free flow writing on websites. The dataset consists of 50 sets of digital handwritten images, including 323 characters. To achieve high-quality input for the OCR process, preprocessing techniques such as converting the image into grayscale, performing basic morphological operations like erode and dilate, and noise removal are implemented. The grayscale image is easier to work with than color images. Morphological operations, such as erosion and dilation, filter the image based on shapes, isolate individual elements, and smooth contours. Noise removal removes unnecessary pixels using median blur operations. The Tesseract engine uses page layout analysis to detect text areas in images, divide them into blobs, define lines, perform word recognition in two passes, and choose the highest confidence word.

Handwritten text recognition (HTR) is gaining popularity in the digital age, but its text decoding is limited by a structured character dictionary. Natural Language Processing (NLP) techniques like Machine Translation and Grammatical Error Correction have shown promising results. This work aims to apply alternative spelling correction techniques in post-processing of an HTR system to achieve competitive results and integrate with any post-processing approach, regardless of the dictionary between both systems.

This paper [8] presents an offline Handwritten Text Recognition (HTR) system and its linguistic restrictions when using the Hidden Markov Model (HMM) with a language model. A Spelling Correction system is proposed as post-processing to the HTR system, enabling experimentation of eight independent spelling correction techniques. The proposed encoder-decoder model, using 1024 hidden units and Luong Attention, achieved the best average error rates, improving sentences by 54%. Future research aims to expand the linguistic model independently.

The paper [9] proposes an automatic method for evaluating Bangla subjective answer scripts using keyword matching and linguistic analysis. The model was tested on 20 question answer scripts, with a minimum relative error of 1.8%. The method is particularly useful in Bangladesh, there are many students, mostly in Bangla and English, making it

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

challenging for teachers to evaluate their skills. The proposed system for evaluating Bangla answer scripts uses a systematic methodology, starting with keyword generation and frequency ratio generation. It then collects answers from online resources like Wikipedia. A comparison algorithm is proposed to score the student's answer based on the frequency ratio of keywords and their frequency ratio. The system also analyzes grammatical and spelling mistakes using a dictionary. A system was developed to evaluate Bangla subjective answer scripts, involving 20 questions and two answers. Teachers scored answers and calculated absolute and relative errors. The system is automated, faster, and uses machine learning algorithms. It's designed for various sentences and handwritten inputs, with a relative error under 10%.

[10] image processing is a technique that converts images into digital form through scanning and applying standard methods. It's a rapidly growing technology with applications in various business sectors. Common methods include digital, optical, and analog. Techniques include enhancement, restoration, and compression.

Image enhancement involves accentuating image features like boundaries or contrast to enhance the usefulness of a graphic display. It includes gray level and contrast manipulation, noise reduction, edge crisping, filtering, interpolation, magnification, and pseudo coloring. Image restoration filters the observed image to minimize degradation effects, focusing on extraction or accentuation of features. Image compression minimizes the number of bits required to represent an image, used in broadcast TV, remote sensing, military communication, and more. The process involves importing the image, analyzing, and manipulating it, and outputting the result.

Document segmentation is a crucial preprocessing step in document image analysis, dividing documents into text and non-text components. This study aims to enhance the document segmentation approach to segment handwritten bill templates, which contain heterogeneous components like image, printed text, handwritten text, and graphical images. The proposed algorithm detects arbitrarily oriented text in images using the enhanced Radical Sector Coding (RSC) algorithm. The algorithm uses a simple method to extract invariant topological features, including the Center of Mass (CoM), scaling invariance, and rotation invariance.

The paper [11] introduces an automatic short answer grading method using Bloom's Taxonomy, WordNet, and Latent Semantic Analysis models, showing high correlation with teacher grading for three high school students. Research on automatic grading in virtual learning environments has gained relevance due to its benefits like low costs and instant feedback. Discursive questions are crucial for evaluating students' reading, interpretation, and writing abilities. The Bloom Taxonomy, a hierarchical classification of learning stages, is used in research on discursive questions. It categorizes cognitive domains into six categories: knowledge, understanding, application, analysis, synthesis, and evaluation. Discursive questions help students understand expectations and study material. The study utilized data from an automatic short answer grading study, utilizing Latent Semantic Analysis (LSA) models and WordNet models. The Apache Jena framework was used to load data, and agreement among evaluators was calculated.

Automatic short answer grading is a field assessing students' natural language answers. This paper [12] presents a new method for automatic short answer grading, incorporating the best of both approaches. The study evaluates automatic grading approaches for students using two public datasets released in 2012 as part of the "SemEval-2013 Task 7: The Joint Student Response Analysis" competition. The datasets consist of questions, student answers, reference answers, and grade labels. The competition data split evaluates system performance in generalization and prediction from unknown questions, with three distinct testssets. The proposed method uses four distinct features to better model tasks, with three used in all test sets and the fourth for the UA test set. Preprocessing is applied to Text Statistics and Semantic Similarity, while Lexical Similarity measures lexical level similarity. Four metrics groups are considered: token-based, edit-based, sequence-based, and compression-based. The study focuses on creating bag-of-ngrams features for Unseen Questions and Unseen Domains scenarios using text statistics and similarity. The bag-of-ngram model was trained using Random Forests and Extreme Gradient Boosting, with parameter tuning for each scenario.

This systematic review [13] explores the automatic short answer grading (ASAG) field, which assesses short natural language responses to questions automatically. This method can provide a deeper assessment of students' knowledge. This paper discusses the use of Computer-Based Assessment (CBA) in teaching, focusing on short answers in natural language and recalling external knowledge. It discusses the benefits of CBA in automating evaluation processes, providing faster feedback, saving time, and allowing teachers to easily track class performance.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

His systematic review uses machine learning to automatically grade short answers in a research question. The review process involves planning, conducting, and reporting, ensuring a fair and comprehensive review. The nature of datasets used in reviewed papers varies greatly, including topics, language, student characteristics, grading scale, average size, and number of questions. Natural Language Processing (NLP) techniques are utilized in the preprocessing stage to model answers for computer interpretation. Over 10 techniques are used, including punctuation, numbers, symbols, acronym expansion, sentence segmentation, case normalization, and tokenization. Other techniques aggregate value at lexical, syntactical, or semantic levels.

Evaluators often issue assessments as part of ongoing learning methods, but human effort is high and costly. Existing computer-based evaluation systems lack scale and support for subjective questions, offline examinations, and dynamic content delivery. Mobile Agents are effective for distributing applications. Automated essay scoring systems, such as Project Essay Grading, Intelligent Essay Assessor, E-rater, and IntelliMetric, have been developed since the 1960s.

This approach [14] proposes a novel method for automatic assessment of descriptive text answers. The system consists of three modules: scanning, training, and scoring. The scanning phase scans documents and extracts student answers, while the training phase uses Natural Language Processing (NLP) to filter out essential parts. The learning phase consists of training, testing, with unscored answers provided as PDF files for awarding marks. The system consists of preprocessing, comparison, and scoring. The training phase uses scored answers and high-weighted answer keys as input. Preprocessing extracts important features of the input, mapping each preprocessed answer and key into vector spaces based on TF_IDF score and cosine similarity score. The system uses NLTK, an NLP Toolkit, to extract semantics from answer sheets. Key steps in preprocessing include grammar check, tokenization, stop word removal, synonym and antonym checking.

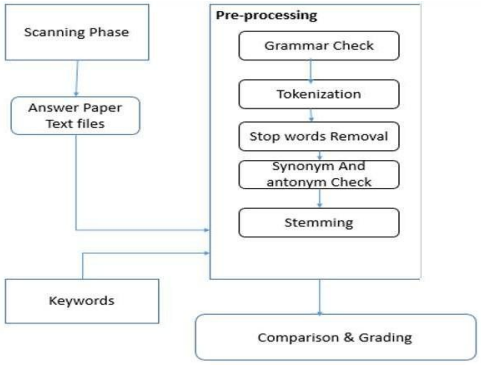

Figure 1:System architecture for proposed system in [14]

The study [15] explores the use of natural language processing and semantic analysis techniques for automatic grading of short answers in teaching-learning, using unsupervised machine learning approaches like latent semantic analysis and WordNet, for improved performance. The research employs the Latent Semantic Analysis (LSA) method and the WordNet Model to measure text similarity. The LSA method verifies word and sentence similarity through contextual use, while the WordNet model organizes nouns, verbs, adverbs, and adjectives. The Shortest Path method, a path-based measure, was chosen for its best results when combined with LSA.This paper presents a research approach to solve linguistic problems using five steps: data collection, pre-processing, similarity scores calculation, linear regression for grade prediction, and performance metrics.

Advances in technology have increased the availability of handwritten documents, making them difficult to search. To address this, document image word spotting techniques have emerged as an alternative to traditional OCR-based methods. These techniques find specific words in document images by comparing features extracted from word images.

Word spotting and word recognition are two methods used to search for specific words in document images. Word spotting involves segmenting documents into words or sub-words, which are then described using features. These features are used to train classifiers and build models or indexes. In word recognition, word images are given to the system, and features are extracted to classify them based on the previously generated models.

Preprocessing in document image word spotting involves binarization, noise removal, skew correction, and line/word segmentation, with algorithms applying these tasks in various orders.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

Word image representation is crucial for document image word spotting, with feature extraction reducing storage requirements and improving performance. Feature indexing uses techniques like latent semantic indexing, locality preserving indexing, and distance-based hashing, impacting searching time.

Handwritten document word spotting faces challenges due to writing styles' larger variability, infinite word count, poor quality, and age-related degradation issues. Despite these challenges, proposed approaches provide efficient support for constrained documents, ensuring efficient retrieval from vocabulary.

This review [16] explores various word spotting techniques, categorizing them based on extracted features. Profile-based features capture word outline, with Manmatha and Croft achieving an average precision rate of 72.65%. Rath and Manmatha found DTW to be better. Kolcz et al. proposed a line-oriented approach for handwritten documents, Sigappi et al.

The paper [17] introduces a joint feature distribution principle for designing novel discriminative features for handwritten document understanding, evaluating seventeen features for writer identification, script recognition, historical manuscript dating, and localization.

The paper explores the use of textural and grapheme-based features in handwritten manuscripts to identify the writer, script, date, and localize the text.

handwritten document analysis, including textural-based and grapheme-based methods. Textural-based features, such as local binary pattern (LBP) and co-occurrence local binary pattern (CoLBP), are widely used in texture recognition and writer identification. LBP is a gray-scale invariant textural feature, while CoLBP is a 2D histogram representing the probability of co-occurrence LBP patterns along ink contours. Run-length histograms (RLH) are used in handwritten document analysis, quantizing the run-lengths of certain patterns along a given direction. LBPruns feature computes the run-lengths of local binary patterns formed with n parallel scanning lines along a given direction with inter-line distance d on binarized images. The number of n determines the number of possible local binary patterns, and the inter-line distance determines the spatial resolution of local binary patterns.

Computer-based software could save time and cost by providing a simple, cost-effective, and reliable OCR approach for multiple-choice questions. It could examine OMR answer sheets, detect feedback, and compare student answers with a database master key. OpenCV is an open-source software tool for computer programmers, enabling candidates to input total choice values into spreadsheets. It [18] uses optical reading technology and can process thousands of answer sheets daily. Microsoft Visual Studio is used for Windows, websites, and web services. The system consists of a title and response answer sheet, with a personnel-identification section for exam candidate identification. It is simple, efficient, and reliable, requiring a low-cost scanner and minimal storage space. The system can be used in education institutes to track student performance, evaluate staff feedback, increase question paper sets, and be implemented at the micro level in government and private sectors for better organization. The system requires a black and white printer for printing.

This [19] paper presents a hybrid Automated Essay Grading System (AEGS) that uses natural language processing and a neural network grading engine to automatically grade student essays.

The Automated Essay Grading System (AEGS) is a system that uses a neural network and writing features analysis to grade students' essays. It compares new essays to pre-graded ones, provides feedback for grammar and stylistic errors, and uses natural language processing tools to identify undesirable features.

AEGS is an automated writing instruction system that identifies five main grammar, usage, and mechanical errors, including agreement, verb formation, wrong word use, missing punctuation, and typographical errors. It also highlights style revisions, such as passive sentences and excessive repetition. The system uses four modules: grammar, usage, mechanics, confusing words, undesirable style, and essay-based discourse elements, providing accurate feedback to improve performance.

This [20] paper reviews Automatic Short Answer Grading (ASAG) research, identifying 35 systems within five temporal themes and six common dimensions. It concludes that an era of evaluation is the newest trend in ASAG research, paving the way for the consolidation of the field, as past efforts have been ad hoc and non-comparable.

The Educational Testing Service (ETS) categorizes research on Automatic Supportive Assessment (ASAG) into active and passive questions, with the depth of learning being the first level. Recall questions require original answers, while

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

automatic grading primarily targets recall questions. The second level includes specific question types like "natural language" and "graphic questions," focusing on text-based questions, given the extensive literature on ASAG.

information extraction techniques extract structured data from unstructured sources like free text for various applications. AutoMark is an automated method for pattern matching on parse tree representations of teacher and student answers for grading. WebLAS identifies important segments of teacher answers, while Thomas '03 addresses ASAG as a boolean pattern matching problem with thesauri support.

The paper [21] tackles challenges in combining OCR with machine translation and keyword spotting in business and government operations, using natural language processing technologies and integrating features from multiple resources.

The paper discusses two challenges in a government-sponsored project for image document analysis and translation. The first involves translating Arabic handwriting OCR output, which requires sentence-level processing. The second challenge is recognition-based keyword spotting, which retrieves recognized speech, image documents, and video text using keywords. A conditional random field-based sentence boundary detector outperforms a rule-based method, and a new method improves search relevance using novel features like OOV name detection.

Keyword search is a method for accessing information from foreign image documents, divided into OCR-based and image feature-based approaches. OCR-based keyword spotting is used due to its performance, preprocessing, and enhanced NLP technologies. Name spotting is a focus, but OCR-based keyword spotting is affected by recognition errors.

[22] Optical Character Recognition (OCR) is a field of research in pattern recognition, artificial intelligence, and machine vision, used in real-life applications. The field is divided into two parts: machine-printed character recognition and handwritten character recognition. Feature extraction is crucial in OCR.

OCR involves acquiring a scanned document image in a specific format, undergoing several stages including data acquisition, pre-processing, segmentation, and feature extraction. These steps include binarization, noise reduction, normalization, skew correction, slant removal, segmentation, and feature extraction. Binarization converts gray-scale images into binary, noise reduction removes isolated specks, normalization reduces data size, segmentation breaks single text lines, words, and characters, and feature extraction represents each character as a single character.

Feature extraction and classification techniques, including Fractal theory, Gabor filters, and SVM, are crucial for character recognition, with SVM classifier achieving 95.04% accuracy and offline k-NN classifier achieving 94.12% accuracy.

The study [23] investigates the use of reading texts for automatic short answer scoring in German language learning, finding textual features improve classification accuracy, while other models suggest less instructor supervision. This paper investigates the role of text in short answer scoring in reading comprehension exercises, comparing student answers to instructor-provided answer-based models. It evaluates various models and presents results from an annotation study. Short answer scoring (SAS) is a method that automatically assigns labels to individual learner answers, either binary or fine-grained. Previous work on SAS, including early systems, has been applied to foreign language learning.

The authors propose an alignment-based model for German data, reimplementing the Meurers model. The model preprocesses material using NLP tools, extracts synonyms and semantic types from GermaNet, and uses spellchecking to identify words not accepted by a German spellchecker. The model uses the German CREG corpus for reading comprehension tasks.

[24] Digital Image Processing and computer vision are advancing the development of optical character recognition systems for various applications, including automated data entry, bank cheque analysis, and handwritten pictogram interpretation. Researchers worldwide have proposed methodologies to solve these problems, with OCR systems being robust for many languages and input modes. Efficient systems may be developed for specific applications.

The field of handwritten document image recognition has a long history, dating back to the early sixties and seventies. Character images were initially modelled as blocks defined geometrically by their edges and edge information. Later, integrated segmentation and interpretation systems emerged, and algorithmic and computational techniques were developed to process images at high levels. The trend shifted towards automated techniques for recognition, incorporating efficient algorithms for pre-processing, feature extraction, and classification. The use of lexicons,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

dictionaries, and language models is also being introduced to achieve higher recognition rates. Handwritten vs. handprinted text is often misinterpreted, but handwritten documents have mono-spaced characters and uniform base-line character images. Off-line and online recognition modes of obtaining input for recognition are also discussed.

The feature extraction phase is crucial in the OCR process, impacting recognition efficiency. Selecting a suitable method ensures feature invariance and high performance. Techniques are available in languages like Assamese, Persian, Thai, Devanagari, Chinese, and Roman.

[25] Handwriting recognition is increasingly popular due to its applications in banking, envelope reading, and document and video recognition. Stochastic modeling, such as Hidden Markov Models (HMMs), is effective for modeling handwriting and line recognition. Combining different HMM-based classifiers is more efficient and complementary, as it focuses on clustering character states and refining them based on hand movement variations.

First HMM system uses the sliding window approach, which includes dynamic information related to derivative features computed from neighboring windows. These features are useful for introducing context from neighboring characters and are highly useful for the recognition of handwritten words. The system begins with normalization and preprocessing, focusing on deslanting images and capturing features independent of word height. Feature extraction is based on El-Hajj et al.'s work, extracting geometric features from each window, including pixel densities, density features, background/foreground transitions, and gravity center position. A subset of features is baseline dependent. The features related to pixels densities are calculated directly on gray-level images, and the features extraction parameters are optimized on a validation database. The system introduces context at the feature extraction level through derivative features, which represent the dynamics of features around the current window. The derivation is computed using a regression, with the first and second-order regressions known as delta and delta-delta coefficients

TABLE I. Table of Comparison Results

Sl no	Author/year	Research /Work Paper	Methodology	Technique	Dataset/Input	Experiment/ Observation
1	Najam *et al.* (2023)	Analysis of Recent Deep Learning Techniques for Arabic Handwritten-Text OCR and Post-OCR Correction	OCR techniques are used to read and recognize the text present in an image.	Deep learning using CNN, RNN,LSTM and CTC	Publicly available data sources were used for this work	The paper analyses the OCR and text correction for Arabic text, offering insights into the best-performing architectures, challenges with datasets, and suggestions for future research
2	Prerana, M. S., et al. (2023)	Eval-Automatic Evaluation of Answer Scripts using Deep Learning and Natural Language Processing	Deep learning and natural language processing model.	CNN, RNN, ResNet, BERT, LSTM	IAM Dataset, Crowd-sourced Handwritten Samples, Pesuacademy ESA Manuscript Samples	The text emphasizes challenges in handwriting recognition, proposes the use of GCP OCR to alleviate training complexities, introduces a unique language processing algorithm for improved accuracy, and advocates for GPT-3 in the context of text

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

						summarization.
3	Riza, L.S., Firdaus, Y., Sukamto, R.A. et al. (2023)	Automatic generation of short-answer questions in reading comprehension using NLP and KNN	NLP method is used to process data in the form of text and KNN is used to choose the best question. based on training data.	NLTK, Elementary Sentence Extracted System	Article taken from the BBC news website and the library use is "newspaper3k".	The evaluation involvesachieving a grammatical correctness rate of 76.19%, and precision rate of 80% and feasibilities.
4	Das, Bidyut, et al. (2022)	Automatic question generation and answer assessment for subjective examination	WSM method for fusion, keyword-based question generation method for analyze the effectiveness,weighted sum method for answer evaluation then compared the proposed method with the state-of-the-art.	MCDM-based fusion technique.	DBSYLLABI dataset,25 model answers for each question to evaluate 100 answers of students in the dataset.	The work introduces an automated question generation system using MCDM-based approaches for single sentence evaluation, with limitations like dataset sensitivity and weight setting
5	Mursari et al. (2021)	The effectiveness of image preprocessing on digital handwritten scripts recognition with the implementation of OCR Tesseract	OCR process involves stages like preprocessing techniques including convert into grayscale,basicmorphological operations.	OCR engine Tesseract 4.x	50 sets of image format digital handwritten image including the total of 323 characters.	Preprocessing techniques significantly improved OCR performance by 79.26%, enhancing image quality and reducing noise, thereby enhancing character recognition accuracy in digital handwritten images
6	Bahel et al. (2021)	Text similarity analysis for evaluation of descriptive answers	NLP tools and techniques including text semantics analysis and Siamese Manhattan LSTM algorithm.	Similarity measure	A dataset comprising of 4 students answer sheets.	The model's evaluation was accurate, but not as precise as manual evaluation. If converted to a percentile-based system, the scores would be similar.
7	Agarwal, Mayank, et al. (2021)	Autoeval:Anlp approach for automatic test evaluation system.	NLP is used to extract text from the response document. The scoring marks for an answer text would be determined using a similarity test. Cosine	Natural language Processing [NLP], NLTK, Similarity Measure	student's written response is given as input.	The proposed system saves 300% of time and is 75-87.5% effective, eliminating human effort and time.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

			similarity, Jaccard similarity, bigram similarity, and synonym similarity			
8	Neto et al. (2020)	Towards the Natural Language Processing as Spelling Correction for Offline Handwritten Text Recognition Systems	Optical Character Recognition (OCR) technologiesused to transcribe the content into the digital environment	artificial neural networks as optical models,Convolutional and Recurrent Neural Network (CRNN).	Bentham [35]; IAM [36]; RIMES [37]; Saint Gall [38]; and Washington [39].	The proposed models show a 54% improvement in correction rate compared to statistical approaches, mainly due to encoder-decoder architecture and attention mechanism.
9	Bhole et al. (2020)	Automatic Grading and Adaptive Question Selection for English Language Testing	Natural Language Processing (NLP) for grading of answers and makes use of a methodology to assess voice-based answers using software application	Deep learning model	UiPath platform	Performance prediction parameters can be calculated, and improvement scope can be defined for each examinee individually.
10	Kaur et al. (2019)	Proposed approach for layout and handwritten character recognition in OCR	RSC algorithm for detection of arbitrarily oriented text in an image	Image processing	Shopkeeper Bills	handwritten and printed characters in different files and enhancing character recognition techniques for complex images like graphs and tables.

III. CONCLUSION

The literature review on smart answer sheet evaluation concludes by pointing to a dynamic and changing environment at the nexus of technology and education. The combined results highlight the many advantages of intelligent answer sheet assessment. Among the benefits mentioned in the research are increased grading efficiency, a decrease in subjective biases, and the ability to manage huge assessment volumes. One key element that enhances learning and promotes a more flexible learning environment is the capacity to give students timely and tailored feedback.

Nonetheless, the review of the literature also discusses the difficulties in implementing intelligent assessment systems. Thoughtful consideration must be given to ethical issues, such as worries about fairness and openness in algorithmic decision-making. Furthermore, a common feature is the requirement for thorough training programs for teachers to enable them to use and incorporate these technologies into current pedagogical frameworks.

The literature review offers a strong starting point for further study and advancement in intelligent response sheet evaluation. The survey's findings provide educators, legislators, and technologists with a road map for working together to successfully negotiate the challenges of integrating intelligent grading systems in a variety of educational contexts. A more sophisticated grasp of the advantages and disadvantages of using intelligent assessment technologies is made possible by the synthesis of current knowledge, which opens the door to further advancements in educational assessment techniques.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

REFERENCES

[1]. Riza, L.S., Firdaus, Y., Sukamto, R.A. et al. Automatic generation of short-answer questions in reading comprehension using NLP and KNN. Multimed Tools Appl 82, 41913–41940 (2023)

[2]. Prerana, M. S., et al. "Eval-Automatic Evaluation of Answer Scripts using Deep Learning and Natural Language Processing." International Journal of Intelligent Systems and Applications in Engineering 11.1 (2023): 316-323.

[3]. ajam, R.; Faizullah, S. Analysis of Recent Deep Learning Techniques for Arabic Handwritten-Text OCR and Post-OCR Correction. Appl. Sci. 2023, 13, 7568. https://doi.org/10.3390/app13137568

[4]. Das, Bidyut, et al. "Automatic question generation and answer assessment for subjective examination." Cognitive Systems Research 72 (2022): 14-22.

[5]. Bahel, Vedant, and Achamma Thomas. "Text similarity analysis for evaluation of descriptive answers." arXiv preprint arXiv:2105.02935 (2021).

[6]. Agarwal, Mayank, et al. "Autoeval: A nlp approach for automatic test evaluation system." 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON). IEEE, 2021.

[7]. Mursari, Lily Rojabiyati, and Antoni Wibowo. "The effectiveness of image preprocessing on digital handwritten scripts recognition with the implementation of OCR Tesseract." Computer Engineering and Applications Journal 10.3 (2021): 177-186.

[8]. Neto, A.F.d.S.; Bezerra, B.L.D.; Toselli, A.H. Towards the Natural Language Processing as Spelling Correction for Offline Handwritten Text Recognition Systems. Appl. Sci. 2020, 10, 7711. https://doi.org/10.3390/app10217711

[9]. Hussain, Md Gulzar, et al. "Assessment of bangla descriptive answer script digitally." 2019 International Conference on Bangla Speech and Language Processing (ICBSLP). IEEE, 2019.

[10]. Kaur, Manpreet, Navdeep Singh Randhawa, and Vishal Garg. "Proposed approach for layout and handwritten character recognization in OCR." Int. Res. J. Eng. Technol 6.4 (2019).

[11]. Haendchen Filho, Aluizio, et al. "Bloom's taxonomy-based approach for assisting formulation and automatic short answer grading." Brazilian symposium on computers in education (Simpósio Brasileiro de Informática na Educação-SBIE). Vol. 29. No. 1. 2018.

[12]. Galhardi, Lucas B., et al. "Exploring distinct features for automatic short answer grading." Anais do XV Encontro Nacional de Inteligência Artificial e Computacional. SBC, 2018.

[13]. Galhardi, Lucas Busatta, and Jacques Duílio Brancher. "Machine learning approach for automatic short answer grading: A systematic review." Advances in Artificial Intelligence-IBERAMIA 2018: 16th Ibero-American Conference on AI, Trujillo, Peru, November 13-16, 2018, Proceedings 16. Springer International Publishing, 2018.

[14]. Sijimol, P. J., and Surekha Mariam Varghese. "Handwritten short answer evaluation system (HSAES)." International Journal of Scientific Research in Science and Technology 4.2 (2018): 1514-1518.

[15]. Nau, Jonathan, Aluizio Haendchen Filho, and Guilherme Passero. "Evaluating semantic analysis methods for short answer grading using linear regression." Sciences 3.2 (2017): 437-450.

[16]. Ahmed, Rashad, Wasfi G. Al-Khatib, and Sabri Mahmoud. "A survey on handwritten documents word spotting." International Journal of Multimedia Information Retrieval 6 (2017): 31-47.

[17]. He, S. and Schomaker, L., 2017. Beyond OCR: Multi-faceted understanding of handwritten document characteristics. Pattern Recognition, 63, pp.321-333.

[18]. Kulkarni, Dhananjay, Ankit Thakur, and Jitendra Kshirsagar. "Automatic Answer Sheet Evaluation Using Efficient & Reliable OCR System." (2017).

[19]. Shehab, Abdulaziz, Mohamed Elhoseny, and Aboul Ella Hassanien. "A hybrid scheme for automated essay grading based on LVQ and NLP techniques." 2016 12th International Computer Engineering Conference (ICENCO). IEEE, 2016.

[20]. Burrows, Steven, Iryna Gurevych, and Benno Stein. "The eras and trends of automatic short answer grading." International journal of artificial intelligence in education 25 (2015): 60-117.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

[21]. Chen, J., Cao, H. and Natarajan, P., 2015. Integrating natural language processing with image document analysis: what we learned from two real-world applications. International Journal on Document Analysis and Recognition (IJDAR), 18, pp.235-247.

[22]. Tawde, Gaurav Y., and Jayshree Kundargi. "An overview of feature extraction techniques in ocr for indian scripts focused on offline handwriting." International Journal of Engineering Research and Applications 3.1 (2013): 919-926.

[23]. Horbach, Andrea, Alexis Palmer, and Manfred Pinkal. "Using the text to evaluate short answers for reading comprehension exercises." Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 1: Proceedings of the Main Conference and the Shared Task: Semantic Textual Similarity. 2013.

[24]. Singh, V., Kumar, B. and Patnaik, T., 2013. Feature extraction techniques for handwritten text in various scripts: a survey. International Journal of Soft Computing and Engineering (IJSCE), 3(1), pp.238-241.

[25]. A. -L. Bianne-Bernard, F. Menasri, R. Al-Hajj Mohamad, C. Mokbel, C. Kermorvant and L. Likforman-Sulem, "Dynamic and Contextual Information in HMM Modeling for Handwritten Word Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 10, pp. 2066-2080, Oct. 2011, doi: 10.1109/TPAMI.2011.22.