Department of Computer Science and Engineering

Bangladesh University of Business and Technology (BUBT)

Student's Id and Name	21225103299-Junaeid Hasan Bijoy
Capstone Project Title	Automated Handwritten Grading System
Supervisor Name & Designation	Ahmed Shafkat, Assistant Professor
Course Teacher's Name & Designation	Dr. Md shafiqul Islam Professor

Aspects	Paper#1 **Aspects of a Standardized Automated System for Screening Children's Handwriting**
Problem Statement	There is no standardized, automated system for screening children's handwriting skills, especially at early educational levels. Current tools lack consistency, rely on manual assessment, and ignore critical position and movement features needed for accurate diagnosis.
Key Contributions	• Introduces the concept of **position features** (e.g., hand, pen, body posture) alongside static and dynamic features in handwriting analysis. • Proposes a **multi-sensory, feature-rich framework** for comprehensive handwriting evaluation. • Discusses integration of **AI and feature fusion** for accurate classification and screening.
Objectives/Goal	To design a **standardized automated handwriting screening system** for children that incorporates dynamic, static, and position-based features to identify early developmental handwriting difficulties and promote standardized evaluation methods.

Methodology/Theory	• Utilizes **three data types**: static (offline script), dynamic (real-time writing data), and position features (posture, grip, etc.). • Emphasizes **feature extraction, fusion, normalization**, and **AI-based classification**. • Incorporates expert decision-making and real-time monitoring for system feedback.
Software Tools/Setup Details	The paper does not specify exact tools, but mentions: • Use of **tablet PCs** and **cameras** for data acquisition. • Technologies: **Signal processing, image/video processing, and AI classification algorithms**.
Test/Experiment	Comparative analysis of existing systems (e.g., TRAZO, MEDDRAW, ComPET) and proposed model structure. Parameters include: • Pen trajectory • Pressure • Speed • Position data (head, hand, paper)
Test Data/Dataset Source	No specific dataset is used in this conceptual paper. It references prior systems and test conditions involving children aged 3–11, especially those with writing challenges (e.g., dysgraphia).
Result Analysis	As a conceptual framework, no empirical results are presented. However, the paper argues the system will: • Improve diagnosis accuracy • Support education and public health • Enable real-time assessment and intervention

Obstacles/Challenges	• Lack of a common handwriting standard • Need for cross-disciplinary collaboration • Difficulty capturing and analyzing position features • Early-stage development of multi-feature fusion and classification
Terminology/Keywords	Handwriting, static/dynamic/position features, fusion, normalization, dysgraphia, AI classification, pen trajectory, early childhood education, automated screening.
Final Summary	The paper proposes a standardized, automated system to screen children's handwriting by integrating static, dynamic, and position features. Recognizing the lack of consistent evaluation tools, it introduces a novel approach to include body posture and hand movement data during writing. The system aims to identify early handwriting difficulties using advanced feature extraction, fusion, and AI-based classification. By combining insights from education, medicine, and engineering, the model emphasizes the importance of multi-sensory data collection and normalization for robust assessment. It also highlights the limitations of existing tools like TRAZO and MEDDRAW, which are either clinical or assistive rather than screening-focused. Although no dataset is used directly, the paper synthesizes findings from prior research and existing platforms. The proposed model outlines a pipeline involving data acquisition through tablets and cameras, preprocessing, feature selection, and expert feedback. Ultimately, the system envisions transforming early education by offering scalable, accurate, and insightful handwriting evaluations.

Department of Computer Science and Engineering

Bangladesh University of Business and Technology (BUBT)

Student's Id and Name	21225103299-Junaeid Hasan Bijoy
Capstone Project Title	Automated Handwritten Grading System
Supervisor Name & Designation	Ahmed Shafkat, Assistant Professor
Course Teacher's Name & Designation	Dr. Md shafiqul Islam Professor

Aspects	Paper#2 **Automatic essay exam scoring system : a systematic literature review**
Problem Statement	There is an increasing demand for efficient and accurate evaluation of essay-type exam answers in remote learning environments, especially post-COVID-19. Manual grading is time-consuming and subjective, making automated essay scoring essential.
Key Contributions	• Conducted a **systematic literature review (SLR)** on 10 relevant studies between 2017–2021. • Identified and categorized the **algorithms and datasets** used in automatic essay scoring. • Highlighted performance metrics like **QWK (Quadratic Weighted Kappa)** and accuracy rates to evaluate systems.
Objectives/Goal	To explore existing research on **automated essay scoring systems**, identify common **methods, tools, and datasets**, and provide insights for future development of reliable, scalable automated assessment tools.

Methodology/Theory	• Followed the **Systematic Literature Review (SLR)** framework (based on Kitchenham). • Defined three research questions: themes, algorithms, and datasets used. • Applied **data extraction**, **inclusion criteria**, and **quality assessment** for analysis.
Software Tools/Setup Details	Various tools and algorithms used in the reviewed studies: • **Machine Learning**: CNN, LSTM, MDLSTM, Naive Bayes • **Statistical Models**: Hidden Markov Models • **Image Processing**: Hough Lines, Horizontal Projection Profile • **NLP Libraries**: NLTK, OpenNLP
Test/Experiment	• Handwritten essays from students in grades 7–10. • Public datasets: Hewlett Foundation AES, Automated Student Assessment Prize (ASAP). • Custom datasets collected via scanned documents or digital pens.
Test Data/Dataset Source	No specific dataset is used in this conceptual paper. It references prior systems and test conditions involving children aged 3–11, especially those with writing challenges (e.g., dysgraphia).
Result Analysis	• **QWK scores** ranged up to 0.88, showing high reliability in systems like MDLSTM and CNN-based models. • Accuracy of some models (e.g., HMM and CNN) reached **up to 100% and 92.86%** respectively in controlled datasets.

	• AES systems are shown to be effective but vary by dataset type and algorithm used.
Obstacles/Challenges	• Variation in essay structure makes universal modeling difficult. • Handwriting recognition still lacks precision, especially for cursive/poor-quality inputs. • Limited availability of standardized and large datasets.
Terminology/Keywords	Automated Essay Scoring (AES), QWK, CNN, LSTM, MDLSTM, NLP, HMM, SLR, handwriting recognition, dataset preprocessing, machine learning classification
Final Summary	This paper provides a systematic review of ten recent studies on automated essay scoring systems, aiming to identify effective methods, tools, and datasets. It highlights the growing need for scalable, accurate grading due to the shift to online education post-COVID-19. The studies reviewed use diverse techniques like CNN, LSTM, Hidden Markov Models, and handwriting recognition, often evaluated with QWK metrics. The datasets include handwritten student responses and public sources like the Hewlett Foundation AES. Results show high accuracy and reliability, with some models achieving over 90% performance. However, challenges remain in handwriting variability, data preprocessing, and lack of standard datasets. The paper concludes that automated systems are effective but must evolve with better feature engineering and evaluation standards. This review offers valuable guidance for researchers aiming to improve AES systems in educational settings.

Department of Computer Science and Engineering

Bangladesh University of Business and Technology (BUBT)

Student's Id and Name	21225103299-Junaeid Hasan Bijoy
Capstone Project Title	Automated Handwritten Grading System
Supervisor Name & Designation	Ahmed Shafkat, Assistant Professor
Course Teacher's Name & Designation	Dr. Md shafiqul Islam Professor

Aspects	Paper#3 **Automated Assessment of Multimodal Answer Sheets in the STEM domain**
Problem Statement	The paper identifies the challenge of automating the grading process for STEM (Science, Technology, Engineering, and Mathematics) assessments, which involve complex, multimodal answer sheets containing both textual responses and diagrams (e.g., flowcharts). Unlike subjects like Literature, STEM assessments require evaluating quantitative analysis, handwritten text, and diagrams, which are difficult to assess due to their variability in representation (e.g., different labels like "yes/no" vs. "true/false" in flowcharts conveying the same meaning). The lack of efficient, accurate, and fair automated grading systems for such multimodal inputs hinders scalability and objectivity in STEM education

	evaluation. The problem is compounded by issues like inaccurate handwriting recognition and the need for contextual understanding in diagram evaluation.
Key Contributions	• Developed an automated system for evaluating multimodal STEM answer sheets, integrating textual and diagrammatic assessments. • Proposed a methodology combining OCR (CRAFT, TrOCR, Azure OCR), object detection (YOLOv5), and LLMs (Mistral AI) for comprehensive evaluation. • Introduced a novel approach to convert diagrams into textual representations for semantic evaluation using LLMs, addressing visual variability in diagrams. • Achieved high accuracy in grading, with 196 out of 250 questions meeting expected marks across five STEM subjects.
Objectives/Goal	The primary goal is to develop an efficient, accurate, and fair automated assessment system for STEM answer sheets that handles both textual answers and diagrams. The system aims to minimize manual intervention while ensuring reliable grading, addressing challenges like handwritten text recognition and contextual diagram evaluation. It seeks to inspire similar efforts for other diagram types and multi-domain evaluations, ultimately enhancing the learning experience and promoting educational equity through technology-driven assessment.
Methodology/Theory	• **CRAFT**: CNN-based model (VGG-16 backbone) for text detection, generating score maps using affinity and region scores. • **TrOCR**: Transformer-based OCR model for handwritten and printed text extraction, trained on IAM Handwriting Database and synthetic data. • **YOLOv5**: Object detection model for identifying diagram components (blocks, arrows), using CSP-Darknet53 backbone,

	Neck, and Head components. • **Azure OCR**: Deep learning-based OCR for extracting steps and mathematical expressions in flowcharts.
Software Tools/Setup Details	The paper does not specify exact tools, but mentions: • Use of **tablet PCs** and **cameras** for data acquisition. • Technologies: **Signal processing, image/video processing, and AI classification algorithms**.
Test/Experiment	The experimental analysis was conducted on unit test papers from five STEM subjects: Data Structures and Algorithms (DSA), Computer Networks (CN), Design and Analysis of Algorithms (DAA), Computer Systems Design and Fundamentals (CSDF), and Software Engineering (SE). Each subject included 10 papers with 50 questions (5 questions per paper).

Test Data/Dataset Source	• **Textual Data**: Handwritten answers from student test papers across five STEM subjects (DSA, CN, DAA, CSDF, SE), with 10 papers per subject, each containing 5 questions (250 questions total).
	• **Diagrammatic Data**: A dataset of handwritten and synthetic flowchart images, including correct and incorrect diagrams. Referenced datasets include:
	  • Montellano et al. (2022): 775 handwritten flowchart images in Spanish and English.
	  • FloCo dataset (Shukla et al., 2023): 11.8K flowchart images with corresponding Python codes.
	  • Bresler et al. (2016): 300 finite automata and 672 flowchart diagrams.
	• **Model Answer Key**: Predefined answers and diagrams for comparison, created for

	each subject.
Result Analysis	The system evaluated 250 questions across five subjects, with 196 questions (78.4%) meeting expected marks: • **DSA**: 39/50 questions correct. • **CN**: 41/50 questions correct. • **DAA**: 36/50 questions correct. • **CSDF**: 38/50 questions correct. • **SE**: 42/50 questions correct.
Obstacles/Challenges	• **OCR Accuracy**: Handwritten text, especially cursive, posed challenges for OCR. • **Diagram Discrepancy**: Visual differences in diagrams (e.g., "yes/no" vs. "true/false" in flowcharts) led to lower scores if compared directly. • **Handwriting Variability**: Non-legible handwriting caused misinterpretation in grading. • **Diagram Quality**: Faintly drawn boundaries or poorly labeled diagrams reduced evaluation accuracy. • **Scalability**: Handling diverse reference answers and scaling the system for larger datasets or other diagram types remains a challenge.
Terminology/Keywords	Automated grading, OCR, diagrams, flowcharts, LLMs, text evaluation, STEM education.
Final Summary	The paper addresses the challenge of automating STEM assessment by developing a system that evaluates multimodal answer sheets containing text and diagrams, tackling issues like handwritten text recognition and diagram variability. Key contributions include a novel methodology integrating CRAFT, TrOCR, YOLOv5, Azure OCR, and Mistral AI to achieve accurate grading, with 78.4% of 250 questions meeting expected marks across five STEM subjects. The system aims to enhance efficiency, fairness, and

	educational equity, with potential expansion to other diagram types and languages. The methodology involves segmenting answer sheets, extracting text and diagram data, and using LLMs for semantic evaluation. Tools like FastAPI and Mistral-7B support a scalable framework. Experimental analysis on student test papers and flowchart datasets validated the system’s effectiveness, though challenges like OCR accuracy and diagram quality persist. This work signifies a milestone in automated assessment, poised to reshape STEM