International Journal of Contemporary Computer Research (IJCCR), Vol.1 Issue.1 (January, 2025)
ISSN: 2600-9048

A Review of Automated Essay Scoring (AES)

Yazeed Al Moaiad[1], Mohammad Alobed[2], Mahmoud Alsakhnini[3], Wafa Al-Haithami[4]

Faculty of Computer & Information Technology,[1,2,3]
Faculty of Finance and Administrative Sciences,[4]
Al-Madinah International University,
Kuala Lumpur, Malaysia

[1]Yazeed Al Moaiad, [2]Mohammad Alobed, [3]Mahmoud Alsakhnini, [4]Wafa Al-Haithami
yazeed.alsayed@mediu.edu.my, shamona.2018@gmail.com, m.alsakhnini@gmail.com, alhaithamiwafa@gmail.com

Received: Sep 15, 2025; Revised: Oct 10, 2025; Accepted: Nov 20, 2025; Published: Jan 03, 2025

***Abstract* - Student achievement is measured via tests. Now, humans assess Manual evaluation gets harder when teacher-to-student ratios grow. Manual evaluation is slow and unreliable. Online examinations replaced pen and paper. Computer-based testing only evaluates multiple-choice questions, not essays and short answers. Many researchers have worked on automated essay grading and short answer scoring for decades, but assessing an essay by considering all elements is challenging. Few studies evaluated content, but many styles. This article discusses essay-grading automation. We explored artificial intelligence and machine learning essay-scoring techniques and research restrictions. Relevance and coherence aren't rated in essays. Automated Essay Scoring (AES) is a difficult undertaking that involves grading student writings. Human inaccuracies, inequity issues, time requirements, and so on are all diminished. Natural language processing, machine learning, deep learning, etc. are just a few of the many methods that may be used for this. High-quality components are essential to the overall efficiency of such systems. This paper's primary objective is to assess various AES tactics in both intra- and inter-domain contexts.**

***Keywords* - Automatic essay scoring; automatic essay grading; semantic analysis; feedback; natural language processing; deep learning;evaluation metrics; transformer models.**

I. Introduction

Assessment is key to teaching. It affects pupils' curriculum comprehension and advancement. Choosing an assessment approach is key to curriculum development. In this change from face-to-face to full-time online instruction, various issues emerged, including how to evaluate student writing online. IT professionals, teachers, testing experts, and politicians all need to work together to improve the reliability of online exams. the author stresses the need for universities to provide a reliable remote proctoring system, instruct instructors in the use of multiple assessment strategies, and inform students of the need of maintaining academic honesty in their coursework [1].

Essays are required on a daily basis. Language and knowledge-based proficiency may be tested via essays. Essays are used to assess language ability in IELTS, GRE, and other examinations. Manually analyzing written essays in today's technologically sophisticated society is time-consuming. It is necessary to grade user essays automatically. People desire to assess and enhance their language abilities. This endeavour will be beneficial to them. This initiative will benefit students, scholars, and anybody else who wants to enhance their writing [2].

Feedback on writing performance is just as important as level measurement. Recognizing strengths/weaknesses and the influence of each characteristic is required for producing adaptive feedback for learners. Feedback enables learners to recognize the gap between their current performance and the desired outcome and self-regulate in order to attain their goals. Feedback is an essential component of learning and has a significant influence on students' progress. Unsatisfactory feedback is a significant problem for college students. Despite the development of digital technology and different teaching and learning methodologies in recent years, paper-based exams have remained the primary method of evaluating learners [3].

In 2012, the Hewlett Foundation ran a competition on Kaggle to find a means to computerize student evaluations. The winner of this contest received the Automated Student Assessment Award (ASAP). Furthermore, quadratic weighted kappa was employed continuously throughout the competition to assess the consistency between human judges' judgments and the outcomes of automated systems. A total of 154 different organizations attempted to make a forecast. The winning team had a very high kappa score of 0.81407. Kaggle offers all of the code and data you may possibly want for your data science projects. With access to over 50,000 publicly available datasets and 400,000 freely downloadable notebooks, you'll be able to complete any study in no time [4, 5]

When evaluating AES systems, it is customary to first compare the ratings generated by the AES to the scores that were determined by human raters. The Pearson correlation, the Spearman correlation, and the QWK correlation are some of the statistical tests of correlation or agreement that may be used for this purpose [6]. The QWK measure was chosen to serve as the official assessment standard for the ASAP project. The QWK is a widely used measurement for determining the degree to which raters agree with one another (a.k.a. inter-rater reliability).

II. Literature Review

In the next part, a comprehensive literature study will be carried out on the most current AES research investigations. Table 1 compares popular AES.

TABLE 1. COMPARES POPULAR AES

Research title
1. A Neural Approach to Automated Essay Scoring [7]
Research Method: Use recurrent neural networks to learn the relationship between an essay and its score.
Results: ○ The system, based on LSM networks, beats a strong baseline by 5.6% in quadratic weighted Kappa without feature engineering.
2. Skip Flow: Incorporating Neural Coherence Features for End-to-End Automatic Text Scoring [8]
Research Method: This study describes a novel neural design that adds neural coherence to vanilla neural network models.
Results: ○ The technique outperforms feature engineering baselines and deep learning models on the ASAP dataset.
3. Attention-based Recurrent Convolutional Neural Network for Automatic Essay Scoring [9]
Research Method: Recurrent Convolutional Neural Network (RCNN) to learn text representation and assess essays automatically
Results: ○ The approach beats state-of-the-art neural networks for autonomous essay grading, according to ASAP data
Future work: ○ Future study addresses neural models for cross-domain AES.
4. Automated Essay Scoring: A Siamese Bidirectional LSTM Neural Network Architecture [10]
Research Method: Subject matter experts provided sample essays to illustrate grading criteria. The writers also develop an essay and example essay input pair. The authors provided a symmetrical neural network AES model that accepts the new input pair. The Siamese Bidirectional Long Short-Term Memory Architecture (SBLSTMA) model captures essay semantics and grading requirements.
Results: ○ The authors examine the ASAP dataset using the SBLSTMA model established for the Automated

Evaluation of Student Work task. The research shows that approach neural network technology is better than others.

- Approach beats baseline by 5%. By deconstructing the model, the authors discover that one with distance input is superior than one without.

Future work:

- Using this strategy, test the hypothesis in several areas.
- Investigate data augmentation to enhance the essay dataset.

5. **Automated essay scoring with string kernels and word embeddings** [11]

Research Method: string kernels and word embeddings, state-of-the-art outcomes in text classification tasks including identifying native languages and Arabic dialects, String kernels compare strings by counting n-grams.

Results:

- The dataset utilized in the study was taken from kaggle.com. All the essays given are already human-graded.
- The authors contrasted in-domain and cross-domain technique on the Automated Student Assessment Prize data set with various state-of-the-art approaches.
- In-domain and cross-domain comparative studies show that string kernels, alone and in conjunction with word embeddings, perform best on essay grading. The authors' report better outcomes with shallow learning than with deep learning.

6. **Automatic Essay Scoring Incorporating Rating Schema via Reinforcement Learning** [12]

Research Method: The authors present a reinforcement learning essay scoring methodology using quadratic weighted kappa as guidance.

Results:

- Experiment findings on benchmark datasets show QWK training is successful.

Future work:

- In the future, reinforcement learning may explore additional scoring activities besides categorization.

7. **Automated Essay Grading using Machine Learning Algorithm** [13]

Research Method: machine learning techniques, Linear regression technique will be utilized for training the model along with making the use of various other classifications and clustering techniques.

Results:

- The dataset utilized in the study was taken from kaggle.com. All the essays given are already human-graded.
- Comparing human and machine-graded essays, it's evident that the computer can judge an essay like a person.

Future work:

- Future scope may be large. Model semantic and syntactic quality.
- Utilize semantic parsers.
- Using neural networks or PBFs, improve linear regression.

8. **The effectiveness of using a hybrid mode of automated writing evaluation system on EFL students' writing** [14]

Research Method:

- The effectiveness of a hybrid automated writing evaluation system on EFL students' writing.
- Learners wrote an essay in MY Access and saved it. Second session, they rewrote essays based on program comments. In hybrid mode, the same students changed their contribution based on the instructor's comments and kept contributing essays through MY Access.

Results:

- AWE programs improve writing production, according to research.

- In the hybrid scenario, students outperformed AWE students.

Future work:

- Small sample size limits the research. Future research should include more people.
- Future study should follow students' activities while exposed to writing using AWE to see how their performance is associated with their interactions with quick feedback and if multiple activities might contribute to optimum writing output.

9. Automated Essay Scoring based on Two-Stage Learning [15]

Research Method: Two-Stage Learning Framework (TSLF) combines feature-engineered and end-to-end AES techniques.

Results:

- Experiments show the efficiency and resilience of Two-Stage Learning Framework (TSLF).
- TSLF exceeds baselines on 5/8 prompts and achieves state-of-the-art average performance without negative samples.
- TSLF beats the features engineered and end-to-end baselines after adding adversarial essays and is quite robust.

10. AAEE – Automated evaluation of students' essays in Arabic language [16]

Research Method: A system based on Rhetorical Structure Theory and Latent Semantic Analysis

Results:

- The AAEE demonstrates 90% of test essays were accurately evaluated and a correlation of 0.756 between automated and instructor grading. This beats the Arabic human-human correlation of 0.709.

Future work:

- In future development, the authors will incorporate more system features to enhance assessment findings.
- The authors plan to utilize word2vec for semantic similarity.
- The authors want to enhance assessment outcomes by collecting more essays and having each one assessed by more than one instructor.

11. Generalizability of Automated Scores of Writing Quality in Grades 3-5 [17]

Research Method: The authors evaluated the dependability of writing assessment and if an automated essay scoring (AES) system may enhance judgements made regarding students in Response to Intervention frameworks for writing (RTI-W).

Results:

- Results showed that accurate low-stakes judgments, such as sending a student to intervention, may be made by averaging scores from a single narrative, informational, and persuasive prompt (three total) or two prompts per genre (six total) if delivered to samples of struggling writers.
- Reliable high-stakes decisions, such as determining a student's eligibility for special education, could be made by averaging scores from two prompts per genre (six total) when administered to samples of students representing the full range of writing ability or 4–5 prompts per genre (12–15 total) if administered to struggling writers.
- These results assist educators and researchers enhance writing assessment based on decision type and student demographic. AES may be a viable alternative to standard human-scored writing exams for use in RTI-W, while construct validity study is required.

12. Automatic Evaluation for Arabic Essays: A Rule-Based System [18]

Research Method: This work presents a rule-based system to automatically analyze Arabic essays based on free textual essay analysis, regardless of predetermined model essays.

Results:

- To test the technique, the authors gathered unconstrained pieces from university-level Arabic authors. This dataset was carefully reviewed using our good-writing rubric. Our algorithm properly scored 73% of essays overall. Moreover, system performance varies per criteria.

Future work:

- This system provides a basis for future AES systems with complicated characteristics and machine learning.

13. Developing an Automated Essay Scorer with Feedback (AESF) for Malaysian University English Test (MUET): A Design-based Research Approach [19]

Research Method: This research builds an essay grading system. AESF enables teachers to learn new marking topics, create tasks, assess progress, and certify scores, while students may practice writing and develop autonomously. At now, the system can grade and offer on-going feedback to users on two well-trained subjects. On untrained topics, AESF can nevertheless provide grammatical criticism.

Results:

- Five iterations led to Automated Essay Scorer with Feedback (AESF). This approach lets instructors gather marked essay examples to train on new essays. The instructor may then assign tasks, monitor student progress, offer comments, and correct scores.
- Students may practice writing essays and get feedback at any step in the process to receive paragraph and overall essay grades.
- 24 instructors from 5 schools tried the system in actual classrooms and liked it.

Future work:

- More students and markers will be recruited to contribute to the corpus' growth, which will benefit everyone involved.

14. Automated Essay Grading [2]

Research Method:

- Machine learning techniques, categorizing a corpus of textual items into a few distinct grades.
- Support-vector-machine produces a score after text pre-processing, feature extraction, topic analysis, and quick similarity evaluation.

Results:

- The project Automated Essay Grading System improves manual and current automated essay grading systems by incorporating Machine Learning and Natural Language Processing methods.

Future work:

- An essay prompt similarity measure improves the model's accuracy.
- Plagiarism detection would be beneficial.
- The method only benefits the user if the authors can identify plagiarism.
- Copying won't boost users' writing talents.

15. Neural Automated Essay Scoring Incorporating Handcrafted Features [20]

Research Method: Authors mix personally produced essay-level qualities with a DNN-AES model.

Results:

- This work utilized the ASAP dataset, a standard in AES research.
- Accuracy is better for short essay prompts than lengthy essay prompts.
- The approach enhances AES accuracy, according to experiments.

Future work:

- Applying the approach to additional DNN-AES models is another goal.

- Adding layers after the feature input layer may enhance accuracy.

16. Enhancing Automated Essay Scoring Performance via Fine-tuning Pre-trained Language Models with Combination of Regression and Ranking [21]

<u>Research Method:</u> The authors use many task failures to fine-tune pre-trained language models to increase AES performance.

<u>Results:</u>

- Mean square error, batch-wise ListNet, and dynamic weights confine the scores.
- Authors utilize Quadratic Weighted Kappa to measure model performance on the Automated Student Assessment Prize dataset.
- The approach beats state-of-the-art neural models by 3% and the latest statistical model. The model outperforms other state-of-the-art models on two-story stimuli.

<u>Future work:</u>

- Future work will include using the complete lengthy text with the pre-trained BERT model.

17. Automated Essay Grading System using NLP Techniques [22]

<u>Research Method:</u> Proposed approach grades essays using two factors. Simple characteristics include spelling, grammar, punctuation, and proportional problems. Complex features include discourse analysis, topic analysis, and writing style identification. Many present AES systems don't examine the semantic aspects of the essay.

<u>Results:</u>

- The suggested system uses Kaggle datasets. Model accuracy and outcomes match instructors' grades.
- The system fails on fresh data, and although 0.73 is nice, it's not acceptable for real-world deployment.

<u>Future work:</u>

- Changes in client needs might strengthen the system.

18. The Application of Deep Learning in Automated Essay Evaluation [23]

<u>Research Method:</u> Neural network-based deep learning is suited for AES research and development since AES focuses on writing quality. Deep learning technology can grade human-scored writings as input and output. Deep learning may be utilized to choose linguistically significant writing attributes for AEE model generation.

<u>Results:</u>

Experiments demonstrate feasibility, thus further investigation is needed.

<u>Future work:</u>

- Deep learning learns probabilistic, not semantic, linguistic properties from a corpus. Deep learning has showed promise in NLP, however there is no ideal theory to describe its concept, hence it lacks theoretical foundation.
- Most deep learning-based NLP research is data-driven, and few studies can be combined with linguistics. Future NLP, linguistics, cognitive science, and other fields should encourage each other to help robots comprehend and analyze language.

19. Automated Essay Grading System Using Deep Learning [24]

<u>Research Method:</u> The project aims to create a system that automatically scores essays and papers. The suggested grading system receives essays as input and scores them using deep learning methods such as LSTM and dense layers

<u>Results:</u>

- The algorithm makes excellent predictions based on word count, sentence count, prevalence, and elements of speech
- By completing this study, the authors have shown the potential of neural networks in processing natural language challenges.

<u>Future work:</u>

- Although we projected essay grades based on prevalence, this study has more potential.
- Better performance and accuracy may be achieved by training the model on bigger, more complex datasets.

20. Automated Essay Scoring System using Multi-Model Machine Learning [25]

Research Method: Using natural language processing methods like Global Vectors for Word Representation (GloVe) and neural networks for photo categorization, the authors built a time- and cost-efficient automated grading system.

Results:

- The authors evaluated our technique using hand rated essays from a 2012 Kaggle competition on automated essay grading.
- The results reveal that the machine accurately scores most essays and grades the remainder similarly to a person.
- The technique is successful in analyzing essay prompts and may be used to aid another grader or as an independent grader.

Future work:

Using Term Frequency - Inverse Document Frequency, the authors intend to counteract lengthier essays. This permits less important words, which occur more often in lengthier texts, to have less weight and the information to be more articulated.

21. A Comparative Study of Pretrained Language Models for Automated Essay Scoring with Adversarial Inputs [26]

Research Method: This research examines three AES models using distinct text embedding methods: Global Vectors for Word Representation (GloVe), Embeddings from Language Models (ELMo), and Bidirectional Encoder Representations from Transformers (BERT)

Results:

- The authors' established a new evaluation measure, robustness, which assesses models' performance against adversarial inputs, and evaluated the three text embedding strategies by their models' resilience. BERT>Glove>ELMo
- This shows BERT is the best paradigm for an AES system with adversarial attacks.

22. Evaluation Toolkit for Robustness Testing of Automatic Essay Scoring Systems [27]

Research Method: The authors offer a model-agnostic adversarial evaluation method and metrics for AES systems to assess natural language comprehension and resilience.

Results:

- The authors discover that AES models are very over stable, thus even substantial alterations (up to 25%) with unrelated material do not lower their score. Unrelated material improves scores, indicating that the models' assessment process and rubrics should be rethought. 200 human raters grade an original and hostile answer to see whether they can tell the difference and if they agree with auto scorers.

Future work:

- Future work involves creating more efficient AES systems utilizing the suggested metrics, merging the metrics to create a more holistic analysis criterion, and upgrading the assessment suite with focus on test type and student education level.

23. Implementing Automated Writing Evaluation in Different Instructional Contexts: A Mixed-Methods Study [28]

Research Method: This research talks about the AWE system MI Write and shows the results of a mixed-methods study that looked at how AWE could be used to teach writing in middle school. The study looked at how AWE could be used in both a traditional process-based approach to teaching writing and a strategy-based approach based on the Self-Regulated Strategy Development model.

Results:

Interview data demonstrated that AWE's impact on teaching was comparable across settings; particularly, the introduction of AWE resulted in both instructional environments taking on purposeful practice features.

Future work:

Future studies should explore the function of writing practice with AWE.

24. A Trait-based Deep Learning Automated Essay Scoring System with Adaptive Feedback [29]

Research Method:

- The authors build a framework that improves the validity and accuracy of a neural-based AES model for evaluating/scoring characteristics.
- The authors expand the approach to deliver trait-specific adaptive feedback depending on essay characteristics.

Results:

- The LSTM-based system beat the baseline by 4.6% in quadratic weighted Kappa (QWK).
- Predicting characteristics scores improves total score prediction.

Future work:

- Future work may emphasize the words and phrases that gave the AES a certain score for additional analysis and adaptive feedbacking, in addition to training and testing the model on a bigger dataset with well-defined rubrics.

25. Automated essay scoring using efficient transformer-based language models [30]

Research Method: Analyze fine-tuned, limited-parameter NLP models using the AES dataset. Ensembling models provide better outcomes with fewer parameters than most transformer-based models.

Results:

- This research purpose was not to reach state-of-the-art, but to demonstrate that meaningful results may be achieved with a small memory footprint and computing budget. The authors' exceeded prior BERT performances using one-third the settings.
- This approach helps categorize and evaluate papers when context is more important than existing transformer-based language models allow.
- Combining techniques that improve performance should lead to smaller, better, and more environmentally friendly models.

26. Integration of Automated Essay Scoring Models Using Item Response Theory [31]

Research Method: The proposed approach uses Item Response Theory (IRT) to produce an average prediction score from many AES models, taking each model's details into consideration to measure examinees' abilities.

Results:

- A latent IRT model gives superior accuracy than individual AES models and averaged scores.

27. Domain Specific Automated Essay Scoring Using Cloud Based Nlp Api [32]

Research Method: The research describes a technique and application framework (PUAnalyzeThis) that uses MeaningCloud API to automatically extract entities, ideas, relations, etc. and generate scores and grades based on their relevance to a subject's graph.

Results:

- The research offers a framework for swiftly assessing a group of texts.

Future work:

- Future research will include enhanced parallelization of text evaluation methods, using the Apache Hadoop cluster in the implementation design. Cluster nodes might leverage thread-level parallelism.
- It's also a good idea to give an extra level of abstraction to guarantee program independence as to which morphological analyzer is utilized.

28. Neural-Network Architecture Approach: An Automated Essay Scoring Using Bayesian Linear Ridge Regression Algorithm [33]

Research Method: The research introduced a neural-network approach architecture and Bayesian linear ridge regression algorithm to increase AES accuracy and reliability.

Results:

- Setting regression parameters and generating the Bayesian Ridge model exhibits data effectiveness. Bayesian Linear Ridge Regression Algorithm fits Automated Essay Scoring.

Future work:

- Future attempts may involve feedback analysis, which is crucial for self-regulated learning.
- The research doesn't incorporate sentiment analysis, which might improve future essay interpretation.
- The study lays the groundwork for future research on automated essay scoring.
- The research will be a milestone in education, particularly essay writing, since it will improve education via higher instructor productivity and speedier student evaluation feedback.

29. Automated Essay Scoring Using Transformer Models [34]

Research Method: This research compares a transformer-based method to a BOW-based logistic regression model.

Results:

- Transformer-based approaches have several benefits, but BOW approaches reduce words to their stems and ignore word order.
- A second advantage of the transformer-based technique, or a language model-based approach in general, is that AES may be improved by switching language models.
- The authors also illustrate how transformer-based models may improve human raters' accuracy and offer instructions on how to use them.

Future work:

- The authors may simply score a task in a different language by switching to a related language model or a multi-lingual model, which enables fine-tuning the same task in several languages concurrently.

30. Efficacy of Deep Neural Embeddings based Semantic Similarity in Automatic Essay Evaluation [5]

Research Method:

- Essay grading using semantic similarity
- The study uses deep neural embedding to compute essay semantic similarity. Traditional text embedding approaches like Jaccard similarity index and TF-IDF estimate semantic similarity. Recent deep neural embedding techniques include ELMo, Google Sentence Encoder (GSE-Lite and GSELarge), and GloVe.

Results:

- Experiments demonstrate semantic similarity is crucial to essay judgment.
- This study gives significant information on the embedding strategy to use in an automated essay assessment.

31. AES Systems Are Both Over stable and Oversensitive: Explaining Why and Proposing Defenses [35]

Research Method: The authors present detection-based protection models that accurately identify oversensitive and over stable samples.

Results:

- The methods accurately identify anomalous attribution patterns and adversarial samples.

Future work:

- The exploratory research hopes to start a dialogue about better modeling autonomous scoring and testing systems, particularly for essay grading.

32. Learning Automated Essay Scoring Models Using Item-Response-Theory-Based Scores to Decrease Effects of Rater Biases [36]

Research Method: This work presents a novel strategy for training AES models using IRT-based ratings to reduce rater bias.

Results:

- Using a real dataset, the authors showed that the suggested technique generates more robust essay scores than standard AES models. Our strategy enhanced rating prediction accuracy.

Future work:

- Future research will assess the suggested approach utilizing other datasets.
- The authors want to repeat these tests using different ASAP essay data.
- This topic's essay score prediction is easier than others with more flexibility or lengthier essays.
- The diversity of essay subjects may increase the rating discrepancies between raters for different topics.
- Developing an end-to-end training approach for the suggested technique is another future research area.

33. Many Hands Make Light Work: Using Essay Traits to Automatically Score Essays [37]

Research Method: The authors offer a technique to score essays holistically using a multi-task learning (MTL) approach, where holistic scoring is the main job and rating essay attributes is the auxiliary task.

Results:

- The BiLSTM system, which is based on multi-task learning (MTL), gives essays good scores on all of their qualities. MTL systems grade essays and attributes 2.30 to 3.70 times faster than single-task learning (STL).

Future work:

- Future work might use characteristic scoring to provide text feedback to the writer, such as identifying where a poor trait score originated.
- The authors also aim to apply this strategy in cross-domain AEG, where we train our system with writings from one question then test it with essays from another prompt.

34. Integrating Cognitive Computing with Machine Learning for Big Data Analysis in Marking Digital Essay Examination [38]

Research Method: This project aims to augment the prior essay-marking application with cognitive-based models and a functional system to prove its practicability.

Results:

- This program will help university professors save paper during exams. This tool may assist West Africa Examination Council cut marking costs, human error, and exam logistics.

Future work:

- This framework implementation in colleges and high schools.

35. Integrating Deep Learning into an Automated Feedback Generation System for Automated Essay Scoring [39]

Research Method: This research contributes the following. It compares three AES algorithms with word-embedding and deep learning models (CNN, LSTM, and BiLSTM). Second, it presents an automatic feedback creation mechanism based on Constrained Metropolis Hastings Sampling (CGMH). Third, it integrates AES and feedback generation into a classifier.

Results:

- The AES algorithm outperforms state-of-the-art scoring models, while the CGMH technique creates semantically-related feedback phrases. Constrained Metropolis Hastings Sampling
- The results support an automated essay grading and

feedback system.

- Implications may lead to models that expose linguistic aspects while attaining high scoring accuracy, as well as feedback corpora to give more semantically-related and sentiment-appropriate feedback.

Future work:

The present work might help create and validate individualized automated feedback for ITSs and other virtual learning systems.

36. The Effects of Data Size on Automated Essay Scoring Engines [40]

Research Method: The authors explore the influence of data amount and quality on three Automated Essay Scoring (AES) engines; Frequency and hand-crafted feature-based model, recurrent neural network model, and pretrained transformer-based language model for categorization.

Results:

- This research shown that LSTMs can perform comparable to transformer-based and BOW-based models in elaboration and organization given adequate data, and that conventions gain more from the pretrained model's starting attributes.

37. Survey Paper on Smart Essay Grader [41]

Research Method: With more individuals taking numerous examinations like the GRE, TOEFL, and IELTS, judging each paper becomes more difficult, as does maintaining a consistent perspective. This project develops a robust interface to help people grade essays. This research helped us extract variables including the Bag of Words, sentence and word counts, average lengths, structure, and organization.

Results:

This approach is good for short datasets.

Future work:

The authors will do all in their power to maintain the most up-to-date version.

38. A novel automated essay scoring approach for reliable higher educational assessments [42]

Research Method: This research provides a transformer-based neural network model for better AES performance utilizing Bi-LSTM and RoBERTa.

Results:

- Comparing essay scoring results with human raters shows that the proposed model surpasses previous techniques in QWK score. Comparative study of findings shows the model's usefulness in automated essay scoring in higher education.

39. The Use of an Automated Writing Evaluation System for Summative Assessment in an EFL Context: The Relationship Between Automated System Scores and Human Raters' Scores [43]

Research Method: Teachers' holistic scores and AWE holistic scores for the same student essays are the focus of this research.

Results:

- There seems to be a connection between a human-rater and an AWE system that is congruent with the correlation between two human-raters.
- AWE's evaluation and instructors' assessments were found to have equivalent scores, score intervals and rationales for their results in each of these three categories.
- Overall, the results may help us better understand how these methods might be used in EFL classrooms to evaluate students' writing abilities.

Future work:

- Future research might examine the usefulness of such systems in giving textual feedback to students and how they can be applied in L2 environments.
- AWE-like technologies may assist create more conducive learning environments and support instructors in constructing more successful writing programs.

40. Automatic Assessment of English CEFR Levels Using BERT Embeddings [44]

Research Method: The authors suggest using neural networks to categorize English written examinations into CEFR levels. The authors use pre-trained Bidirectional Encoder Representations from Transformers (BERT) models to process language efficiently and quickly using attention-based processes and long-range sequence characteristics.

Results:

- The suggested technique is examined on two open-source datasets: EFCAMDAT and Cambridge Learner Corpus for the First Certificate in English (CLC-FCE).
- The experimental findings suggest that the proposed technique may accurately estimate the learner's competency level, especially with large tagged corpora.
- Adding modifications to the supplied text improves automated language evaluation.

41. Countering the Influence of Essay Length in Neural Essay Scoring [45]

Research Method: a basic neural network is introduced to compare essays with varying grades for content similarity

Results:

Dataset features should be taken into account while developing neural essay grading systems, according to the results.

42. Automated Short-Answer Grading using Semantic Similarity based on Word Embedding [46]

Research Method: The mechanism measured the learner's accuracy using word embedding and syntactic analysis.

Results:

- The semantic similarity technique had a 0.70 correlation and 0.70 MAE with the grading reference.

Future work:

- The authors want to enhance word2vec by adding new text corpora as input.
- The authors want to increase the research problem, particularly for short essay replies needing a chain of solutions.

43. Deep learning-based approach for Arabic open domain question answering [47]

Research Method: Arabic Open-domain question answering (OpenQA) uses deep learning. The approach comprises of document retrieval to obtain key paragraphs from Wikipedia and an answer reader to extract the exact response. For passage retrieval, the model uses dense passage retriever and for reading comprehension, AraELECTRA.

Results:

- The dense passage retriever surpasses the classic TF-IDF information retriever in top-20 passage retrieval accuracy and enhances our end-to-end question answering system in two Arabic benchmark datasets.

Future work:

- For future work, the retriever can be improved by combining Dense Passage Retrieval (DPR) with BM25 or other IR models using a hybrid approach.

44. Improving Performance of Automated Essay Scoring by Using Back-Translation Essays and Adjusted Scores [48]

Research Method: In this work, the authors developed a strategy to enhance the number of essay-score pairings utilizing back translation and score modification.

Results:

- The authors analyzed enhanced data using past models. A model employing long short-term memory, frequently used for automated essay assessment, was also utilized to evaluate performance. Augmented data boosted performance.

Future work:

- The authors will develop more efficient, theoretically theoretical, and practical score adjustment algorithms for back-translation essays.

- Also, the strategy will be applied to additional datasets.
- The authors will also study alternative data augmentation approaches and score modification strategies for AES.

45. Survey on Automated Short Answer Grading with Deep Learning: from Word Embeddings to Transformers [49]

Research Method: Recent advances in Natural Language Processing and Machine Learning have affected Automated Short Answer Grading, which we review. The authors augment prior surveys by analyzing contemporary deep learning algorithms. We analyze the move from hand-engineered features to representation learning techniques, which learn representative features from vast data corpora.

Results:

- Methods that mix hand-engineered features with semantic descriptions from the newest models, such transformer architectures, produce the highest results.

Future work:

- The supplied benchmark data sets are limited and not typical of queries and brief reference responses in diverse fields. This impedes the generalization of learning-based approaches, which might overfit.

46. Towards Trustworthy Auto Grading of Short, Multi-lingual, Multi-type Answers [50]

Research Method: This research includes 10 million question-response pairs from several languages encompassing math and language, with considerable variance in question-and-answer syntax. Fine-tuning transformer models for auto grading complicated datasets is beneficial.

Results:

- The top hyperparameter-tuned model has an accuracy of 86.5%, equivalent to state-of-the-art models adjusted to a single question, subject, and language.
- The authors also illustrate how instructors may limit the system's faults and check that the auto grader's performance on specific examinations is near to predicted.

Future work:

- In future work, we want to enhance models by using additional information such as student response times, incorporating feedback to students (e.g., tailored explanations of auto grader choices and human-to-AI coaches), addressing trust difficulties, and detecting dishonest conduct.

47. On the Use of BERT for Automated Essay Scoring: Joint Learning of Multi-Scale Essay Representation [51]

Research Method: The authors present a jointly-learnable multi-scale BERT essay representation. Multiple losses and out-of-domain transfer learning boost performance.

Results:

- Experiment findings reveal that our strategy benefits from cooperative learning of multi-scale essay representation and achieves virtually the state-of-the-art outcome in the ASAP test.
- The technique outperforms all deep learning models in the ASAP challenge for lengthy text problems.

Future work:

- Introducing linguistic information at a suitable scale may enhance segmentation.

48. Enhanced hybrid neural network for automated essay scoring [52]

Research Method: This research presents a hybrid neural network for automated essay grading that extracts and merges linguistic, semantic, and structural essay features.

Results:

- When compared to four state-of-the-art models employing eight public data sets, our approach increases the Kappa index by 1.4%.

49. A systematic review of automated writing evaluation systems [53]

Research Method: The authors reviewed empirical AWE studies to better understand its validity. Using Scopus, we found 105 publications on AWE scoring systems and coded them for argument-based validation

Results:

- AWE scoring research had a rising trend but was heterogeneous in terms of language environments, ecological settings, and educational level;
- a disproportionate number of studies were carried out on each validity inference, with the evaluation inference receiving the most research attention and the domain description inference being neglected;
- most studies adopted quantitative methods and yielded positive results.

50. Automated assessment of subjective assignments: A hybrid approach [54]

Research Method: This research enhances subjective assignment score prediction. Four ML methods are studied using language characteristics. The 3 Layer Neural Network with feature selection worked well with a QWK of 0.678. A novel hybrid model (LF-BiLSTM-att-FS) combines a higher-level DNN with chosen characteristics to add deep learning. Pre-trained glove embedding adds text context.

Results:

- The suggested model improved accuracy, with a QWK of 0.768.

51. Intelligent Scoring of English Composition by Machine Learning from the Perspective of Natural Language Processing [55]

Research Method: This work presented an AI-based (machine learning) automated scoring model for NLP (NLP). With the n-gram model and decision tree (DT) algorithm as technical support, a composition scoring system is built using the composition's content as the extraction feature. Model training uses random forest algorithm.

Results:

- The suggested composition scoring system with N-gram model and DT algorithm and trained by random forest algorithm has great performance, can be used to automatically score students' English composition in schools, and offers a platform for educational applications of ML under AI.

Future work:

- The model's grammar judgement requires improvement. Next, neural networks and other technologies will be used to enhance grammar score.

52. A Survey on Automatic Essay Evaluation System using Machine Learning [56]

Research Method: This research reviews related studies for automated essay scoring.

Results:

- Semantic similarity scores and sentimental analysis are often used. NLTK and POS are mainly utilized.
- Performance classification uses SVM, Naive Bayes, Random Forest, etc.
- They utilized KAPPA measurements (QWK). This document also includes essay evaluation advantages and downsides

53. The use of semantic similarity tools in automated content scoring of fact-based essays written by EFL learners [57]

Research Method: This research examined open-source semantic similarity algorithms for evaluating fact-based essays by EFL learners. A native expert created a gold standard using 50 fact-based writing examples from a Japanese university's academic English course. InferSent, spaCy, DKPro, ADW, SEMILAR, and Latent Semantic Analysis created semantic similarity ratings between student and expert writing samples.

Results:

- o Three course professors evaluated student work manually. To verify human grade validity, samples with discrepant agreement were eliminated and inter-rater reliability was tested using quadratic weighted kappa. After the remaining samples' grades were validated, a Pearson correlation study between semantic similarity scores and human grades indicated that InferSent was the best method for predicting human grades.

54. Application of an Automated Essay Scoring engine to English writing assessment using Many-Facet Rasch Measurement [58]

Research Method: The authors investigated the relationship between scores assigned by an Automated Essay Scoring (AES) system, the Intelligent Essay Assessor (IEA), and grades allocated by trained, professional human raters to English essay writing using two novel procedures: the logistic transformation of AES raw scores into hierarchically ordered grades, and the co-calibration of all essay scoring data in a single Rasch measurement framework. 589 US students in Grades 4, 6, 8, 10, and 12 wrote 3453 essays in response to 18 NAEP writing challenges (4, 8, & 12).

Results:

- o The authors calculated rater, prompt, student, and rubric impacts using Many-Facet Rasch Measurement (MFRM). Within a single Rasch measuring scale, we compared human rater and IEA student ratings. The AES computer matched human assessments and was more consistent.

55. AI-based automated writing evaluation for online language learning: Perceptions of distance learners [59]

Research Method: This research examines adult distance English learners' AWE tool experiences after a four-week writing assignment. Learners' evaluations of the process, how feedback helped their writing, and recommendations for using AWE in distant language learning were acquired using an online open-ended questionnaire.

Results:

- o The data will add to the research on the usefulness of AWE technologies in remote language learning procedures.

56. An effective approach for Arabic document classification using machine learning [60]

Research Method: In this study, an Arabic text classification model was created utilizing Multinomial Nave Bayesian (MNB), Bernoulli Nave Bayesian (BNB), Stochastic Gradient Descent (SGD), Logistic Regression (LR), Support vector classifier (SVC), Linear SVC, and convolutional neural networks (CNN). Al Khaleej data is used for these algorithms.

Results:

- o CNN with character level model outperforms other models in tests.
- o CNN's 98 accuracy beats the best machine learning technology. The offered strategies will be beneficial in social media.

Future work:

- o Future studies might include gathering additional data and using transfer learning models, although data augmentation for Arabic remains an issue due to imbalanced data.
- o Increasing text classification accuracy with hybrid ensembling is another intriguing Arabic text classification field.

57. Deep Learning Architecture for Automatic Essay Scoring [61]

Research Method: The authors present a new architecture based on recurrent and convolution neural networks (CNN). In the proposed architecture, the multichannel convolutional layer learns and captures contextual word n-gram features from word embedding vectors and important semantic notions via max-pooling.

Results:

- o The experiment used eight Kaggle AES datasets. Our suggested approach provides greater grading accuracy than previous deep learning-based AES systems and state-of-the-art AES systems.

Future work:

- Future research will include attention processes and examine their implications on score prediction.

58. An automated essay scoring system: a systematic literature review [62]

Research Method: This article reviews automated essay grading methods. The authors researched AI and ML strategies used to score essays automatically and assessed existing research constraints.

Results:

- The authors noticed that essays aren't graded on relevance and coherence.

59. EssayGAN: Essay Data Augmentation Based on Generative Adversarial Networks for Automated Essay Scoring [63]

Research Method: EssayGAN is a computer program that uses generative adversarial networks to automatically generate essays (GANs)

Results:

- Data augmentation utilizing augmented essays improves the performance of AES systems, according to experimental findings.
- EssayGAN is able to create essays that not only include many sentences but also maintain the coherence of those sentences inside the essay.

60. Ablation Study on Feature Group Importance for Automated Essay Scoring [64]

Research Method: In our work, the authors performed ablation research to find the weakest designed characteristics. The authors employed the same feature engineering and classification methodology as the Automated Student Assessment Prize winners (ASAP).

Results:

- The findings reveal that the prompt is the least influential semantic component. This was due to the classification mode being overfit.

Future work:

- The authors may then investigate the new prompt function.

61. Automatic scoring of Arabic essays over three linguistic levels [65]

Research Method: Essays are graded using lexical, syntactic, and semantic aspects. Sentence structure determines syntactic level. The ultimate essay score is a composite of each level's score.

Results:

- The investigations indicate that trained models reach accuracies and quadratic weighted kappa values comparable to two human raters.
- The findings show that a decision support Arabic scoring system is possible with reasonable assumptions.

62. IBM Watson Studio for Building an Automated Essay Grading System [66]

Research Method: This research compares contemporary essay grading methods based on technique, performance, and targeted traits. LSTM, Watson, RNN

Results:

- Several Explainable AI algorithms were recently disclosed, but no study has examined their role. Automated essay scoring systems grade essay characteristics. Style, substance, and semantic make up these characteristics. Most suggested systems address style, substance, or both.

Future work:

- Future essay grading systems must concentrate on semantic elements; additional semantic attributes may aid with accuracy. Current grading methods can't determine essay accuracy or consistency.

63. Antecedents of Student Character in Higher Education: The role of the Automated Short Essay Scoring (ASES) digital technology-based assessment model [67]

Research Method: This survey-based, cross-sectional investigation was conducted. 412 Indonesian higher education personnel received the questionnaire via random selection. 61.29 percent of respondents were useful. Surveys gathered the data.

Results:

- The research found that digital assessment methods and trust impact student character.
- Character also improves academic success. Student character's moderating function is supported.

Future work:

- Future studies may also examine culture's moderating influence in the framework.
- Culture affects a student's values and views. It's also cross-sectional. Future research may be longitudinal.
- This quantitative research concludes. Future study should use mixed methods for better outcomes.

64. An Automated Writing Evaluation System for Supporting Self-monitored Revising [68]

Research Method: This study describes the creation and assessment of an automated writing evaluation system that merges NLP and user interface design to help students self-revise.

Results:

- A classroom deployment suggests that NLP can properly evaluate where and what sort of modifications students make throughout paper versions, that students participate in self-monitored rewriting, and that the interfaces for displaying NLP findings are effective.

III. Conclusion

The use of automated essay grading systems faces significant obstacles for researchers. Despite the difficulty of developing a reliable AES system, a number of academics are diligently pursuing this goal. Not every evaluation technique is assessed according to the five criteria of consistency, usefulness, fullness of information, feedback, and expert understanding. The majority of essay scoring systems rely on the Kaggle ASAP (2012) dataset, which contains generic student essays and does not demand domain expertise; hence, there is a need for domain-specific essay datasets for training and testing purposes. Over the past few years, a lot of different automated essay scoring (AES) methods have been made. Recent developments in deep learning have shown that using neural network techniques with AES systems may give cutting-edge results. In the past few years, everyone has agreed that using pre-trained models is the best way to solve the vast majority of NLP problems., despite the fact that no system provides students with feedback on their answers.

References

[1] F. Handayani and H. Syarif, "Online Writing Assessment during the Covid-19 Pandemic: Challenges, Options and Scenarios," Proceeding of International Conference on Language Pedagogy (ICOLP), pp. 64-73, 2021.

[2] M. Genty and S. . V, "Automated Essay Grading," Journal of Information and Computational Science, pp. 1456-1462, 2019.

[3] H. W. Lee and Y. M. Cha, "The Effects of Digital Written Feedback on Paper-based Tests for College Students," The Asia-Pacific Education Researcher, p. 489–497, 2022.

[4] Kaggle, "Develop an automated scoring algorithm for student-written essays," 2012. [Online]. Available: https://www.kaggle.com/c/asap-aes.

[5] H. Manik, M. Prasenjit, P. Raman and G. Manish, "Efficacy of Deep Neural Embeddings based Semantic Similarity in Automatic Essay Evaluation," International Journal of Computing and Digital Systems, pp. 1379-1389, 2021.

[6] S. Vanbelle, "A New Interpretation of the Weighted Kappa Coefficients," Psychometrika, p. 399–410, 2016.

[7] T. Kaveh and T. N. Hwee, "A Neural Approach to Automated Essay Scoring," Association for Computational Linguistics, p. 1882–1891, 2016.

[8] Y. Tay, M. C. Phan, L. A. Tuan and S. C. Hui, "SkipFlow: Incorporating Neural Coherence Features for End-to-End Automatic Text Scoring," arXiv, p. arxiv.1711.04981, 2017.

[9] F. Dong, Y. Zhang and J. Yang, "Attention-based Recurrent Convolutional Neural Network for Automatic Essay Scoring," Association for

Computational Linguistics, pp. 153-162, 2017.
[10] G. Liang, B.-W. On, D. Jeong, H.-C. Kim and G. S. Choi, "Automated Essay Scoring: A Siamese Bidirectional LSTM Neural Network Architecture," Symmetry, p. 682, 2018.
[11] M. Cozma, A. Butnaru and R. T. Ionescu, "Automated essay scoring with string kernels and word embeddings," arXiv, pp. 503--509, 2018.
[12] Y. Wang, Z. Wei, Y. Zhou and X. Huang, "Automatic Essay Scoring Incorporating Rating Schema via Reinforcement Learning," Association for Computational Linguistics, pp. 791--797, 2018.
[13] V. V. Ramalingam, A. Pandian, P. Chetry and H. Nigam, "Automated Essay Grading using Machine Learning Algorithm," Journal of Physics: Conference Series, pp. 12-30, 2018.
[14] M. Mohsen, "The effectiveness of using a hybrid mode of automated writing evaluation system on Efl students' writing," Teaching English with Technology, pp. 118-131, 2019.
[15] J. Liu, Y. Xu and Y. Zhu, "Automated Essay Scoring based on Two-Stage Learning," arXiv, pp. 01-07, 2019.
[16] A. M. Azmi and M. F. A.-J. a. M. Hussain, "AAEE – Automated evaluation of students' essays in Arabic language," Information Processing & Management, pp. 1736-1752, 2019.
[17] J. Wilson, D. Chen, M. Sandbank and M. Hebert, "Generalizability of Automated Scores of Writing Quality in Grades 3-5," Journal of Educational Psychology, pp. 619-640, 2019.
[18] A. Alqahtani and A. Alsaif, "Automatic Evaluation for Arabic Essays: A Rule-Based System," IEEE, pp. 01-07, 2019.
[19] S. Y. Ng, C. H. Bong, K. S. Hong and N. K. Lee, "Developing an Automated Essay Scorer with Feedback (AESF) for Malaysian University English Test (MUET): A Design-based Research Approach.," Pertanika Journal of Social Sciences & Humanities, pp. 1451-1468, 2019.
[20] M. Uto, Y. Xie and M. Ueno, "Neural Automated Essay Scoring Incorporating Handcrafted Features," International Committee on Computational Linguistics, pp. 6077--6088, 2020.
[21] Y. Ruosong, C. Jiannong, W. Zhiyuan, W. Youzheng and H. Xiaodong, "Enhancing Automated Essay Scoring Performance via Fine-tuning Pre-trained Language Models with Combination of Regression and Ranking," Association for Computational Linguistics, p. 1560–1569, 2020.
[22] P. Shetty, K. Yadav and P. Kunder, "Automated Essay Grading System using NLP Techniques," International Journal of Engineering and Advanced Technology (IJEAT), pp. 1033-1042, 2020.
[23] S. Ge and X. Chen, "The Application of Deep Learning in Automated Essay Evaluation," Springer-Verlag, p. 310–318, 2019.
[24] H. P. P. V., H. G., K. Ch.Abhigna and N. V. K., "AUTOMATED ESSAY GRADING SYSTEM USING DEEP LEARNING," International Research Journal of Engineering and Technology (IRJET), pp. 952-954, 2020.
[25] W. Zhu and Y. Sun, "Automated Essay Scoring System using Multi-Model Machine Learning," Computer Science & Information Technology (CS & IT), pp. 109-117, 2020.
[26] P. Wangkriangkri, C. Viboonlarp, A. T. Rutherford and E. Chuangsuwanich, "A Comparative Study of Pretrained Language Models for Automated Essay Scoring with Adversarial Inputs," IEEE , pp. 875-880, 2020.
[27] A. Kabra, M. Bhatia, Y. Kumar, J. J. Li and R. R. Shah, "Evaluation Toolkit For Robustness Testing Of Automatic Essay Scoring Systems," arXiv, p. ARXIV.2007.06796, 2020.
[28] C. Palermo and J. Wilson, "Implementing Automated Writing Evaluation in Different Instructional Contexts: A Mixed-Methods Study," Journal of Writing Research, pp. 63-108, 2020.
[29] A. H. Mohamed, A. H. Hesham and N. Mohammad, "A Trait-based Deep Learning Automated Essay Scoring System with Adaptive Feedback," International Journal of Advanced Computer Science and Applications, pp. 287-293, 2020.
[30] C. Ormerod, A. Malhotra and A. Jafari, "Automated essay scoring using efficient transformer-based language models," arXiv, pp. 01-11, 2021.
[31] I. Aomi, E. Tsutsumi, M. Uto and M. Ueno, "Integration of Automated Essay Scoring Models Using Item Response Theory," Springer-Verlag, pp. 54-59, 2021.
[32] G. Pashev, G. Silvia and H. Yuri, "DOMAIN SPECIFIC AUTOMATED ESSAY SCORING USING CLOUD BASED NLP API," International Journal of Computer Science and Mobile Computing, p. 33–39, 2021.
[33] D. Ancheta, M. Magsael, J. J. J. Catulay and J. Costales, "Neural-Network Architecture Approach: An Automated Essay Scoring Using Bayesian Linear Ridge Regression Algorithm," IEEE, pp. 196-200, 2021.
[34] L. Sabrina, M. Christian, H. Christopher, E. Kerstin and B. Steffen, "Automated Essay Scoring Using Transformer Models," Psych, pp. 897--915, 2021.
[35] Y. K. Singla, S. Parekh, S. Singh, J. J. Li, R. R. Shah and C. Chen, "AES Systems Are Both Overstable And Oversensitive: Explaining Why And Proposing Defenses," arXiv, pp. 01-29, 2021.
[36] M. Uto and M. Okano, "Learning Automated Essay Scoring Models Using Item-Response-Theory-Based Scores to Decrease Effects of Rater Biases," IEEE Transactions on Learning Technologies, pp. 763-776, 2021.
[37] R. Kumar, S. Mathias, S. Saha and P. Bhattacharyya, "Many Hands Make Light Work: Using Essay Traits to Automatically Score Essays," arXiv, p. ARXIV.2102.00781, 2021.
[38] A. Oyebola, D. A. Olubukola, A. Wumi, M. Yaw, O. Ayokunle, A. Adesoji, S. Olawale, A. Taiwo and A. Samuel, "Integrating Cognitive Computing With Machine Learning For Big Data Analysis In Marking Digital Essay Examination.," International Journal Of Engineering Research In Computer Science And Engineering, pp. 29 - 32, 2021.
[39] C. Lu and M. Cutumisu, "Integrating Deep Learning into an Automated Feedback Generation System for Automated Essay Scoring," International Educational Data Mining Society, pp. 573-579, 2021.
[40] C. Ormerod, A. Jafari, S. Lottridge, M. Patel, A. Harris and P. van Wamelen, "The effects of data size on Automated Essay Scoring engines," arXiv, p. arxiv.2108.13275, 2021.
[41] S. Nidhi, S. Aman, U. Vaibhav and S. Adhoksh, "SURVEY PAPER ON SMART ESSAY GRADER," International Journal of Engineering Applied Sciences and Technology, pp. 124-127, 2021.

[42] M. Beseiso, O. A. Alzubi and H. Rashaideh, "A novel automated essay scoring approach for reliable higher educational assessments," Journal of Computing in Higher Education, p. 727–746, 2021.
[43] Y. Hilal and İ. K. G. Safiye, "The Use of an Automated Writing Evaluation System for Summative Assessment in an EFL Context: The Relationship Between Automated System Scores and Human Raters' Scores," 4th International Conference on Research in Teaching and Education, pp. 143-153, 2021.
[44] J. S. Veronica and B. Alessio, "Automatic Assessment of English CEFR Levels Using BERT Embeddings," CEUR Workshop Proceedings, pp. 01-07, 2021.
[45] S. Jeon and M. Strube, "Countering the Influence of Essay Length in Neural Essay Scoring," Association for Computational Linguistics, pp. 32-38, 2021.
[46] F. L. Fetty, M. Mutaqin, P. Atina, W. Dana, S. Tri, A. A. Arry and R. Yusep, "Automated Short-Answer Grading using Semantic Similarity based on Word Embedding," International Journal of Technology, pp. 571-581, 2021.
[47] A. Kholoud, J. Amani and A. Areej, "Deep learning-based approach for Arabic open domain question answering," PeerJ. Computer science, pp. 01-21, 2022.
[48] Y.-J. Jong, Y.-J. Kim and O.-C. Ri, "Improving Performance of Automated Essay Scoring by Using Back-Translation Essays and Adjusted Scores," Mathematical Problems in Engineering, pp. 01-10, 2022.
[49] S. Haller, A. Aldea, C. Seifert and N. Strisciuglio, "Survey on Automated Short Answer Grading with Deep Learning: from Word Embeddings to Transformers," arXiv, pp. 01-29, 2022.
[50] S. Johannes, R. Robin and R. Micha, "Towards Trustworthy AutoGrading of Short, Multi-lingual, Multi-type Answers," International Journal of Artificial Intelligence in Education, pp. 01-17, 2022.
[51] W. Yongjie, W. Chuan, L. Ruobing and L. Hui-Ching, "On the Use of BERT for Automated Essay Scoring: Joint Learning of Multi-Scale Essay Representation," ArXiv, p. arXiv:2205.03835, 2022.
[52] X. Li, H. Yang, S. Hu, J. Geng, K. Lin and Y. Li, "Enhanced hybrid neural network for automated essay scoring," Expert Systems, p. e13068, 2022.
[53] S. Huawei and V. Aryadoust, "A systematic review of automated writing evaluation systems," Education and Information Technologies, pp. https://doi.org/10.1007/s10639-022-11200-7, 2022.
[54] N. Birla, M. Kumar Jain and A. Panwar, "Automated assessment of subjective assignments: A hybrid approach," Expert Systems with Applications, p. 117315, 2022.
[55] D. Zhang and X. Yuan, "Intelligent Scoring of English Composition by Machine Learning from the Perspective of Natural Language Processing," Mathematical Problems in Engineering, pp. 01-09, 2022.
[56] N. Patel and P. P. Gamit, "A Survey on Automatic Essay Evaluation System using Machine Learning," International Journal of Scientific Research in Science, Engineering and Technology, pp. 160-166, 2022.
[57] Q. Wang, "The use of semantic similarity tools in automated content scoring of fact-based essays written by EFL learners," Education and Information Technologies, pp. https://doi.org/10.1007/s10639-022-11179-1, 2022.
[58] K. Chan, T. Bond and Z. Yan, "Application of an Automated Essay Scoring engine to English writing assessment using Many-Facet Rasch Measurement," Language Testing, p. 026553222210760, 2022.
[59] A. Taşkıran, "AI-based automated writing evaluation for online language learning: Perceptions of distance learners," Kocaeli Üniversitesi Eğitim Dergisi, pp. 111 - 129, 2022.
[60] Y. M. Abdullah, K. G. Hemantha, H. J., B. J.V. Bibal, M. N. Mourya, C. Channabasava, P. M. and B. R., "An effective approach for Arabic document classification using machine learning," Global Transitions Proceedings, pp. 267-271, 2022.
[61] T. M. Tashu, C. K. Maurya and T. Horvath, "Deep Learning Architecture for Automatic Essay Scoring," arXiv, p. arxiv.2206.08232, 2022.
[62] D. Ramesh and R. D, "An automated essay scoring systems: a systematic literature review," Artificial intelligence review, pp. 2495-2527, 2022.
[63] Y.-H. Park, Y.-S. Choi, C.-Y. Park and K.-J. Lee, "EssayGAN: Essay Data Augmentation Based on Generative Adversarial Networks for Automated Essay Scoring," Applied Sciences, pp. 01-16, 2022.
[64] S. T. JIH and K. T. IAN, "ABLATION STUDY ON FEATURE GROUP IMPORTANCE FOR AUTOMATED ESSAY SCORING," Jurnal Teknologi Maklumat dan Multimedia Asia-Pasifik , pp. 90-101, 2022.
[65] W. Alsanie, M. I. Alkanhal, M. Alhamadi and A. O. Alqabbany, "Automatic scoring of arabic essays over three linguistic levels," Progress in Artificial Intelligence, pp. 01-13, 2022.
[66] Al Moaiad, Yazeed, Zainab Abu Bakar, Ahamd Mhd Emad Diab, Norshuhani Zamin, and Yazrina Yahya. "Cloud service provider cost for online university: amazon web services versus oracle cloud infrastructure." In International Visual Informatics Conference, pp. 302-313. Singapore: Springer Nature Singapore, 2023.
[67] Al Moaiad, Y., Bakar, Z. A., Diab, A. M. E., Zamin, N., & Yahya, Y. (2023, October). Cloud service provider cost for online university: amazon web services versus oracle cloud infrastructure. In International Visual Informatics Conference (pp. 302-313). Singapore: Springer Nature Singapore.
[68] Al Moaiad, Y., Alobed, M., Alsakhnini, M., & Momani, A. M. (2024). Challenges in natural Arabic language processing. Edelweiss Applied Science and Technology, 8(6), 4700-4705.
[69] Al Moaiad, Y., D., Alkhateeb, M., & Alokla, M. (2024). Deep Analysis of Iris Print and Fingerprint for Detecting Drug Addicts. J. Electrical Systems, 20(3), 5639-5644.
[70] Al Moaiad, Y., D., Alkhateeb, M., & Alokla, M. (2024). Enhancing Cybersecurity Practices in Nigerian Government Institutions an Analysis and Framework. J. Electrical Systems, 20(3), 5964-5967.