# Recent Trends in Answer Script Evaluation – A Literature Survey

Maya A.K.R[1,*] Nazura Javed[2] Muralidhara B. L[3]

[1] *Bangalore University, Bengaluru, India*
[2] *St. Francis College, Koramangala, Bengaluru, India*
[3] *Department of Computer Science and Applications, Bangalore University, Bengaluru, India*
[*]*Corresponding author. Email:* mayasanjay@gmail.com

## ABSTRACT

Assessment of answer scripts is an integral part of an examination and education system. A fair, consistent, unbiased, and correct valuation ensures the integrity of an examination system and is important for all education institutions. Since manual valuation is cumbersome and can be biased or influenced by the perception/mood of the evaluator, automatic grading of scripts has become very relevant. Automatic short answer grading (ASAG) techniques have been widely researched in the last decade and have assumed increased relevance because of online teaching and examinations during the Covid-19 pandemic. This review paper focuses on the recent works in the area of automatic answer grading and compares the techniques, methodologies employed, and the consequent results to evaluate their effectiveness. It discusses the advantages and limitations of the techniques by systematically categorizing the questions into both long/short as well as open-ended/close-ended questions and suggests a new model for improving the grading outcomes.

*Keywords: Answer Script Grading, Machine Learning (ML), Natural Language Processing (NLP).*

## 1. INTRODUCTION

Assessing students to measure their progress and ensuring that the curriculum goals are met is a key aspect of the education and learning process. Hence various examinations with different types or categories of questions are conducted for assessing the knowledge and skills of students. Grading of student answer scripts is a time consuming process; the imposed stringent deadlines also make it error prone. Manual grading of student answer scripts may be biased, inconsistent as it is dependent on the skill set, mood swings and perception of the evaluator. Hence an automated model that overcomes the drawbacks of manual assessment and standardizes the grading process, making it more efficient and accurate is essential.

Assessment questions mainly fall under the categories of MCQ, fill-in-the-blanks, match-the-following, essay/short answers (closed-ended), long answers/descriptive (open-ended) and, technical answers [1]. Many earlier works on grading assessment focused on writing skills rather than content [2, 3]. The recent works however, focus on concept/content of the answer and use Natural Language Processing (NLP),

Machine Learning (ML) and hybrid techniques [4]. The importance of content and semantic grading of answer scripts is also highlighted [5].

Most of the works in the area of automatic assessment pertain to grading of short answers. ASAG is a term used for Automatic short-answer grading and involves automatic grading of short answers (typically a few words or a few lines or a paragraph) [1]. Short answer questions are commonly used in the examinations to assess the basic knowledge of the students. Since these answers comprise of a limited number of words, it is easy to compare them with the reference answers and provide the scores/grades effectively [6]. Though short answers questions are mostly close-ended with limited number of matching correct answers, some such questions require the students to express their comprehension and perspectives. Matching techniques for comparison of student answer with reference answer may not yield accurate scores/grades for such open-ended questions [7]. A more intelligent grading system based on machine learning and domain-specific knowledge base would result in enhanced valuation. In the recent years the focus of research has shifted to grading of

descriptive/subjective questions of an open-ended nature [8-11].

Our literature review reveals that research in automatic grading mainly falls under two categories: a) grading essays and b) grading short/long answers. While essays are long (typically spanning two-three pages) with valuation based on content as well as structure/grammar, the short answers are valued primarily on the basis of keywords and concept terms. In this paper, we have reviewed papers published in the recent years from 2016 to 2020 that focus on grading of answers which are both, close ended and open-ended. We find that most of the recent research works for automatic grading use NLP techniques with augmented feature extraction and machine learning/deep learning techniques and restrict their scope to grading of short answers only. In this work we have focused our review efforts on following two aspects:

1) Study of the techniques and methodologies used in the recent works for automatic grading of student answers.

2) A critical evaluation of the above and suggesting a model for an optimum valuation system for both open-ended and close-ended questions.

This paper is structured as follows. Section 2 examines the related works in the area of short answer grading with an emphasis on recent techniques and methodologies. Section 3 describes the taxonomy of the techniques employed in the various literary works under review. Section 4 systematically compares and evaluates the research papers with respect to the datasets, techniques and results. The discussion regarding the techniques employed, their advantages/limitations and guidelines for an optimum grading are contained in section 5. Section 6 concludes the paper by stating possible areas of future research.

## 2. RELATED WORKS - LITERATURE REVIEW

Researching and devising better techniques in education assessment by means of effective tools for candidate evaluation has been the vision of ETS (Educational Testing Service). ETS is guided by its mission "*to help advance quality and equity in education by providing fair and valid assessments, research, and related services. Our products and services measure knowledge and skills . . . for all people worldwide.*" Most of the recent research works also aim to provide efficient automated tools for assessment of short answers, essays and descriptive questions.

The techniques used for the evaluation of answer scripts mainly fall under two categories: a) Similarity techniques based on NLP and b) Machine learning techniques using training datasets. NLP tools for example, NLTK or Stanford NLP tools provide API for tokenization, POS tagging, stop word removal, stemming, lemmatization and summarization. While the tokenization process breaks the text into smaller chunks or tokens, POS tagging identifies the parts of speech for the tokens. Stop words that do not add much meaning to the sentence can be removed during pre-processing and stemming/lemmatization can be used to reduce inflection from words. The NLP-based research works generally use semantic matching techniques for comparison of student answer with reference answer(s) for short answer grading. These works are found mostly effective for close-ended short answer questions [12-16].

Supervised machine learning techniques like classification, regression and deep learning techniques are used for automated answer grading. Bag-of-words model using tf/tf-idf and word embedding models like Word2Vec, GloVe are generally used for representing the text in the corpus. While the Bag-of-words model is sparse, the word embedding is dense. It captures the context, meaning and interrelation of words and hence provides a semantic representation. Though the supervised learning techniques are found to be effective for both close-ended and open-ended questions, their accuracy is good only when large training datasets are available [7, 17-19].

Most of the recent works in ASAG focus on semantic similarity. NLP based text extraction and summarization techniques with augmented similarity measures was found to yield better results for M5 and M10 questions [12]. In this paper, the text is extracted, summarized and then scored on the basis of the five weighted parameters viz. cosine, jaccard, bigram, synonym similarity and grammatical spelling error. The experiments were conducted using three types of questions viz. M5 (5 marks), M10 (10 marks), and M15 (15 marks) questions. It was observed that the difference between the manually scored marks and the automated scores was more for M15 descriptive answers. The main limitation of this work is that the parameter weights used for grading are not computed automatically; they are assigned manually on the basis of a survey.

A supervised machine learning model containing augmented text similarity features that consider the importance of domain-specific words is proposed [17]. The reference answer and the student answer are represented by the two vectors viz. R and S respectively. These vectors generated using off-the-shelf word embedding provide a semantic representation and hence semantic similarity is computed between vectors R and S using cosine similarity measures. This work uses question demoting for removing question words from both student response and reference answer. Moreover, this paper does term weighting based on a variant of tf-

idf. In this variant, idf alone is used for arbitrary words and tf measure based on the frequency of word in internet resources like Wikipedia, is used for domain-specific words. This model was evaluated on two tasks – a) Mohler *et al.* (2011) task, where a real-valued score on a scale of 0 to 5 is computed and b) SemEval-2013 Task where categorical labels viz. correct, partially correct/incomplete, contradictory, irrelevant, and non-domain are assigned. This paper emphasizes the relevance of domain-specific terms in grading; however, this tf-idf variant was not found very effective in case of very short answers. This was especially so when the keywords cannot be distinctly identified. For example: common words like "last", "added", "first", and "removed" are significant in the answer of the definition of stack. Hence, additional term weighting mechanism to identify significant answer words is vital.

Combined and composite models using question, answer-based features, and student models (more than one student model answer) to build feature space are explored [13, 14]. NLP techniques like bag-of-words and similarity matching sometimes fail to determine the correctness/incorrectness of the short answers. The reason for this is the limited amount of text content in short answers is inadequate for extracting sufficient lexical features necessary for grading. The student models can thus play an important role and help to add to the feature space. The answer-based features include length difference, max-matched idf, cosine similarity, weighted text similarity and LSA (Latent Semantic Analysis) defined by measuring the text similarity between student answer and the correct answer(s). Five classic ML algorithms like NB (Naïve Bayes), LR (Logistic Regression), DT (Decision Tree), ANN, and SVM and DBN (Deep Belief Networks) were compared in the experiments. DBN models outperformed the other classic ML models.

Evaluation of descriptive answers is done by extracting features and computing similarity scores by matching the questions and answers [18]. A systematic process involving stages like question classification, answer classification, features extraction and answer evaluation is used. Naïve Bayes classifier is used to classify both, - the question and the answer. Questions are classified as factual, inductive and analytical and the answer classification helps to determine whether the answer matches the question type and the requirements. The model answer is compared with the student answer to get the similarity score. Answers are evaluated for correctness by verifying keywords, synonyms, and the context of the words. The sum of the number of matching keywords and the similarity score value decide the final score.

An intelligent platform that considers aspects like question type, keywords, structural, conceptual and language parameters was designed [19]. This paper classifies the questions into three categories viz. conceptual, specific and analytical. It scores the answers on the basis of similarity score, language score, fuzzy string-matching score, and a concept graph comparing the concepts in the reference answer with that in the student answer. Additionally, this system generates analytical reports like section-wise performance, progress analysis, class-wise performance, exam-wise performance analysis in the form of charts.

The usage of WordNet graphs to measure the text similarity between the student answer and the expected answer for grading a dataset of 400 student answer sheets in social studies have yielded promising results [20]. The semantic relations hypernym, hyponym, holonym and meronym are considered for generating WordNet graphs. The counts of matching nodes are used for generating scores. This work however did not give good grading results for answers containing technical words and misspelled words.

Student answers largely contain free text. Hence deep semantic understanding is required for a good ASAG system. The paper [15] proposes an ASAG framework for grading open-ended answers by modelling semantic relations between student answers and reference answers. The framework comprises of four layers. The Transformer layer transforms the student answers and reference answers into word embeddings. The Multiattention layer captures the relation between the student and reference answers by using additive, subtractive, multiplicative and dot product attention mechanisms. The Inside aggregation layer aggregates the output of the multiway attention layer and, the Prediction layer uses the feedforward layer with softmax function to obtain the output vector. Experiments were conducted with 120,000 student answers. This model outperformed baselines like Logistic regression Gradient boosted decision tree, TextCNN, Multiway Attention Network and Manhattan LSTM (Long short-term memory).

Another deep learning model focusing on open-ended answers uses features extracted from human evaluated answer scripts and the answer key to train the LSTM-RNN model (Long short-term memory networks- Recurrent neural networks) [21]. This model uses glove vector embedding for answer representation providing a semantic approach to grading. The main drawback of this paper is insufficient data used for testing this approach.

Word weights that emphasize the important words/phrases and the knowledgebase (KB) comprising of standard answer are used to grade the student answers using deep learning LSTM networks [16]. It uses three modules viz. input representation and embedding module, knowledge base selection module and the representation fusion module. The input representation layer is the element-wise product of the KB -based

weight matrix, tf-idf weight matrix and the word encoding matrix. The embedding module generates embedding for the KB and questions and stores them in two matrices (memory vectors). The knowledge base selection module uses these two memory vectors and retrieves the relevant information. The representation fusion module merges the embedding of the student answer and the selected facts from the memory vectors and applies this to the Bi-LSTM network to predict the output. The overall accuracy rate of this system is 83.37% and can therefore be applied to assist manual grading [22-26].

Some short answers require students to express their subjective opinions. An ASAG model that integrates both domain-general and domain-specific information to grade semi-open-ended questions is explored [27-29]. This paper uses both domain-general and domain-specific information for feature engineering. This is done by following a two-step training process. The first step trains the CBOW algorithm using Wikipedia [30-34]. The second step trains the CBOW with student answers containing domain-specific terms. The combined word-vector dictionary generated from the two steps of CBOW training is used to represent the student answers. This representation is used as an input to the LSTM classifier which outputs a probabilistic scoring vector. This model shows an enhanced performance on 7 reading comprehension questions and over 16,000 short-answer samples [35].

## 3. TAXONOMY OF RESEARCH

Figure 1 depicts a hierarchical view of the various automatic grading techniques employed in the recent literary works under our review. The first level of organization is concerned with the research techniques

adopted. NLP techniques are used in some of the works for pre-processing and efficient feature extraction. Text similarity computations go beyond keyword matching methods using cosine, bigram and jaccard measures. They include augmented techniques based on semantic similarities using WordNet and concept graphs, and fuzzy matching techniques. The recent papers using ML approach for ASAG mostly use the LSTM-RNN for training the models. These networks use dense word embedding for answer/feature representation and thus provide a semantic dimension to answer grading.
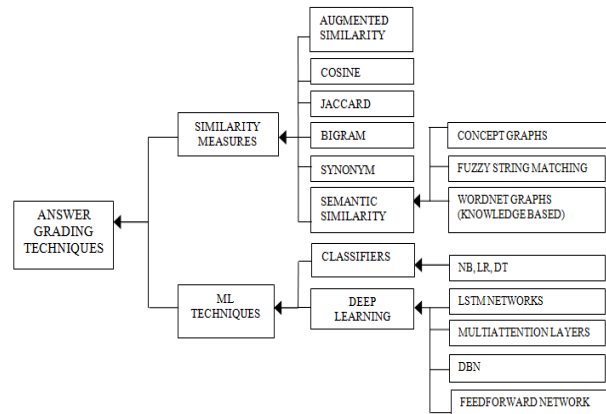


**Figure 1** A hierarchical view of the various automatic grading techniques applied

## 4. A SYSTEMATIC REVIEW

Based on a comprehensive literature review of recent works in automatic answer grading, this section presents a systematic evaluation of the different research works, the techniques and methodologies adopted and results/outcomes thereof.

**Table 1.** Systematic Review

| Sl. No | Research Work/ Paper | Author/ Year | Methodology | Technique | Dataset/ Input | Experiments/ Observations | Remarks |
|---|---|---|---|---|---|---|---|
| 1 | NLP-based Automatic Answer Script Evaluation | Rahman and Siddiqui (2018) [12] | NLP techniques are used for summarization and extraction. The scores are awarded on the basis of similarity measures computed using cosine, jaccard, bigram, and synonyms. | Similarity measures | A dataset comprising of 30 sample questions and answers. The dataset comprises of three types of question answers M5, M10, and M15 | Automatic evaluation of answer scripts was found very useful and often the assigned marks were the same as manually scored marks. | The parameter weights used for grading are not computed automatically; they are assigned manually on the basis of a survey. This is the main drawback. |
| 2. | Fast and Easy Short Answer Grading with High | Sultan *et al.*(2016) [17] | Word embedding used for semantic representation of the student answer and the reference answer. | Semantic Similarity measure | 1. The Mohler et al. task : 80 undergraduate Data Structures questions and | The model shows enhanced performance over | Additional term weighting mechanism is required to |

| | | | | | benchmarks like tf-idf, Lesk, Mohler et al. | identify important answer words. |
|---|---|---|---|---|---|---|
| | Accuracy | | Semantic similarity with augmented similarity measures are used. | | 2,273 student responses. 2. SemEval-2013 datasets (SCIENTSBANK corpus) | |
| 3. | Deep Learning + Student Modeling + Clustering: A Recipe for Effective Automatic Short Answer Grading | Zhang *et al.* (2016) [14] | Combined and composite models using question, answer-based features, and student models are used to build feature space. Six classifiers viz NB, LR, DT, ANN, SVM and DBN are used and their results are compared. | ML algorithms and DBN | Cordillera dataset; 16228 short answers selected from a total of 27868 dialogues | Five classic ML algorithms like NB, LR, DT, ANN, SVM and Deep Belief Networks (DBN) are compared in the experiments. DBN models outperforms the other classic ML models. | DBN outperforms the NB, LR, DT, ANN, SVM classifiers on most parameters like accuracy, AUC, Precision and F-measure (except Recall). |
| 4. | Automatic assessment of descriptive answers in online examination system using semantic relational features | Nandini, Uma Maheswari (2020) [18] | A systematic process involving stages like question classification, answer classification, features extraction and answer evaluation is used. The sum of the number of matching keywords and the similarity score value decide the final score. | Similarity-based measures | 50 questions with10 students answers for each question from a computer science course | The implementation outperforms the other grading system by achieving 95% precision, 94% recall and 94.5% sensitivity. | |
| 5. | An Intelligent System for Evaluation of Descriptive Answers | Bagaria *et al.* (2020) [19] | Uses concept graphs, fuzzy string matching, grammar checking and other similarity metrics. | Grammar checking and similarity measures | A dataset comprising of conceptual, specific, and analytical questions and expected answers from different websites. | The difference between the marks given by manual evaluation and proposed automatic evaluation was less for conceptual and specific type questions and greater for analytical questions. | Answers containing non-textual content like equations, diagrams and tables cannot be scored through this work. |
| 6. | A machine learning approach for automated evaluation of short answers using text similarity based on WordNet graphs | Vij *et al.* (2020) [20] | Uses WordNet graphs for structural and semantic dependencies. | Similarity measure | A dataset comprising of 400 Social Studies subject answer sheets. | RMSE was found to be 0.319 | Fails in evaluating answer scripts with technical words and misspelled words. |
| 7. | Automatic Short Answer Grading via Multiway Attention | Liu *et al.* (2019) [15] | Layered framework with a Multiattention layer that captures the relation between student and reference answers. | Deep Learning model | K-12 dataset. | Outperforms baselines like Logistic regression, Gradient boosted decision | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Networks | | | | tree, TextCNN, Bi-transformer,Ma LSTM and Manhattan LSTM | | |
| 8. | Grading descriptive answer scripts using deep learning | George *et al.* (2019) [21] | Uses glove vector representation and sequential layered framework. | Deep Learning using LSTM-RNN | A dataset from 50 distinct answer scripts based on two different questions. | | Training dataset is insufficient. |
| 9. | Deep Automated Text Scoring Model Based on Memory Network | Yang (2020) [16] | Word weights to emphasize important words/phrases and KB comprising of standard answer are used to grade the student answers using deep learning LSTM networks. | Deep Learning LSTM network | 63 short technical question from the domain of Software Engineering. Each question has 100 scored answers which are used for training. | The overall accuracy rate of this system is 83.37% | *tf-idf* may not be adequate for determining concentration. Single basic concept questions worked better than flow description questions. |
| 10. | An automatic short-answer grading model for semi-open-ended questions | Zhang *et al.* (2019) [7] | CBOW algorithm is used for feature engineering from both domain-general and domain-specific sources. LSTM classifier outputs a probabilistic scoring vector | Deep Learning using LSTM networks | 7 reading comprehension questions and over 16,000 short-answer samples. | | Training the grading model requires many graded student answers for each question. |

## 5. DISCUSSION

The recent years have seen an increased research in the area of automated answer grading, especially in automatic grading of short answers (ASAG). This research has become more relevant because of the Covid-19 pandemic and widespread online classes and online examinations. From our literature review, we found that the works in automatic grading used either a) similarity measures that use both basic and augmented feature extraction for similarity scoring and semantic similarity or b) machine learning and deep learning networks like LSTM.

Our literature study revealed that most of the works in automated grading use non-standard, self-generated datasets. Experiments conducted with standard datasets make the models more comparable. Evaluation of techniques and methodologies becomes concrete and insightful. It was found that though standard datasets like Mohler et al. and SemEval datasets were used, their use was limited and most of the works used datasets generated from internal sources like school/college/university level assessments.

We find that similarity measures (distance-based algorithms) for comparing student answers with reference answer give good results for close-ended short answers. However, these techniques do not give optimum results for open-ended questions/answers that contain less keywords and where students are required to express their own perspectives. ML techniques, combined with KB exhibit enhanced performance for such questions. We believe that categorizing the questions into short/long, open/close-ended would facilitate a targeted approach to automatic grading. We recommend that summarization techniques be applied for open-ended, long questions during the pre-processing stage thus eliminating repeating/similar content.

All the keywords or concepts do not have the same relevance in an answer. Hence weighting of terms on the basis of pre-defined criteria or on the basis of domain-specific KB can help to enhance grading performance. Our study reveals that weighting techniques need further exploration.

## 6. CONCLUSION & FUTURE WORK

Automatic grading of answers scripts makes the valuation system unbiased and consistent. Since the results of the assessments greatly influence the future of the students, a model that overcomes the drawbacks and enhances the grading performance is essential. This review paper studies and compares the recent papers in

the area of automatic grading of answer scripts. Based on our study we conclude that there are two main techniques of answer grading- similarity measures and ML techniques. Similarity based measures do not require a large training set; however, these techniques do not yield good results for open-ended answers. On the other hand, ML techniques enhance the scope of grading systems by giving good results for semi-open-ended questions. The main drawback of using ML techniques is that each question requires a large labelled training set which may be cumbersome and impractical.

Our future work proposes a two-step process for grading. In the first step we propose the use of NLP techniques for chunking or subject-object-predicate extraction from the student and reference answers with a word embedding representation for both. This enables semantic comparison and enhanced scoring. In the second step we envisage a feedback module that generates a labelled training set comprising distinct answers from the dataset used in the first step. This dataset would be used for training the deep learning network for optimized grading outcomes.

## REFERENCES

[1] Burrows, Steven, Iryna Gurevych, and Benno Stein. "The eras and trends of automatic short answer grading." International Journal of Artificial Intelligence in Education 25, no. 1 (2015): 60-117.

[2] Burstein, Jill, Claudia Leacock, and Richard Swartz. "Automated evaluation of essays and short answers." (2001).

[3] Shermis, Mark D., and Jill C. Burstein, eds. Automated essay scoring: A cross-disciplinary perspective. Routledge, 2003.

[4] Shehab, Abdulaziz, Mohamed Elhoseny, and Aboul Ella Hassanien. "A hybrid scheme for automated essay grading based on LVQ and NLP techniques." In 2016 12th International Computer Engineering Conference (ICENCO), pp. 65-70. IEEE, 2016.

[5] Leacock, Claudia, and Martin Chodorow. "C-rater: Automated scoring of short-answer questions." Computers and the Humanities 37, no. 4 (2003): 389-405.

[6] Rokade, Amit, Bhushan Patil, Sana Rajani, Surabhi Revandkar, and Rajashree Shedge. "Automated grading system using natural language processing." In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 1123-1127. IEEE, 2018.

[7] Zhang, Lishan, Yuwei Huang, Xi Yang, Shengquan Yu, and Fuzhen Zhuang. "An automatic short-answer grading model for semi-open-ended questions." Interactive Learning Environments (2019): 1-14.

[8] Patil, Piyush, Sachin Patil, Vaibhav Miniyar, and Amol Bandal. "Subjective answer evaluation using machine learning." International Journal of Pure and Applied Mathematics 118, no. 24 (2018): 1-13.

[9] Prasain, Binita, and Simi Kamini Bajaj. "Analysis of Algorithms in Automated Marking in Education: A Proposed Hybrid Algorithm." In 2020 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA), pp. 1-10. IEEE, 2020.

[10] Surya, K., Ekansh Gayakwad, and M. K. Nallakaruppan. "Deep learning for short answer scoring." Int. J. Recent. Technol. Eng.(IJRTE) 7, no. 6 (2019).

[11] Patil, Rudragouda G., and Syed Zakir Ali. "Approaches For Automation In Assisting Evaluator For Grading Of Answer Scripts: A Survey." In 2018 4th International Conference on Computing Communication and Automation (ICCCA), pp. 1-6. IEEE, 2018.

[12] Rahman, M., and Fazlul Hasan Siddiqui. "Nlp-based automatic answer script evaluation." vol 4 (2018): 35-42.

[13] Zhang, Yuan, Chen Lin, and Min Chi. "Going deeper: Automatic short-answer grading by combining student and question models." User Modeling and User-Adapted Interaction 30, no. 1 (2020): 51-80.

[14] Zhang, Yuan, Rajat Shah, and Min Chi. "Deep Learning+ Student Modeling+ Clustering: A Recipe for Effective Automatic Short Answer Grading." International Educational Data Mining Society (2016).

[15] Liu, Tiaoqiao, Wenbiao Ding, Zhiwei Wang, Jiliang Tang, Gale Yan Huang, and Zitao Liu. "Automatic short answer grading via multiway attention networks." In International conference on artificial intelligence in education, pp. 169-173. Springer, Cham, 2019.

[16] Yang, Shiyan. "Deep Automated Text Scoring Model Based on Memory Network." In 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL), pp. 480-484. IEEE, 2020.

[17] Sultan, Md Arafat, Cristobal Salazar, and Tamara Sumner. "Fast and easy short answer grading with

high accuracy." In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1070-1075. 2016.

[18] Nandini, V., and P. Uma Maheswari. "Automatic assessment of descriptive answers in online examination system using semantic relational features." The Journal of Supercomputing 76, no. 6 (2020): 4430-4448.

[19] Bagaria, Vinal, Mohit Badve, Manasi Beldar, and Sunil Ghane. "An Intelligent System for Evaluation of Descriptive Answers." In 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), pp. 19-24. IEEE, 2020.

[20] Vij, Sonakshi, Devendra Tayal, and Amita Jain. "A machine learning approach for automated evaluation of short answers using text similarity based on WordNet graphs." Wireless Personal Communications 111, no. 2 (2020): 1271-1282.

[21] George, Neethu, P. J. Sijimol, and Surekha Mariam Varghese. "Grading descriptive answer scripts using deep learning." International Journal of Innovative Technology and Exploring Engineering (IJITEE) 8, no. 5 (2019).

[22] N. Shi, L. Tan, W. Li, X. Qi, K. Yu, "A Blockchain-Empowered AAA Scheme in the Large-Scale HetNet", Digital Communications andNetworks, https://doi.org/10.1016/j.dcan.2020.10.002.

[23] Y. Sun, J. Liu, K. Yu, M. Alazab, K. Lin, "PMRSS: Privacy-preserving Medical Record Searching Scheme for Intelligent Diagnosis in IoT Healthcare", IEEE Transactions on Industrial Informatics, doi: 10.1109/TII.2021.3070544.

[24] Z. Guo, L. Tang, T. Guo, K. Yu, M. Alazab, A. Shalaginov, "Deep Graph Neural Network-based Spammer Detection Under the Perspective of Heterogeneous Cyberspace", Future Generation Computer Systems, https://doi.org/10.1016/j.future.2020.11.028.

[25] Rajendrakumar, Shiny, and V. K. Parvati. "Automation of irrigation system through embedded computing technology." In *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, pp. 289-293. 2019.

[26] Naeem, Muhammad Ali, Tu N. Nguyen, Rashid Ali, Korhan Cengiz, Yahui Meng, and Tahir Khurshaid. "Hybrid Cache Management in IoT-based Named Data Networking." *IEEE Internet of Things Journal* (2021).

[27] Le, Ngoc Tuyen, Jing-Wein Wang, Duc Huy Le, Chih-Chiang Wang, and Tu N. Nguyen. "Fingerprint enhancement based on tensor of wavelet subbands for classification." *IEEE Access* 8 (2020): 6602-6615.

[28] Do, Dinh-Thuan, Tu Anh Le, Tu N. Nguyen, Xingwang Li, and Khaled M. Rabie. "Joint impacts of imperfect CSI and imperfect SIC in cognitive radio-assisted NOMA-V2X communications." *IEEE Access* 8 (2020): 128629-128645.

[29] Kumar, M. Keerthi, B. D. Parameshachari, S. Prabu, and Silvia liberata Ullo. "Comparative Analysis to Identify Efficient Technique for Interfacing BCI System." In *IOP Conference Series: Materials Science and Engineering*, vol. 925, no. 1, p. 012062. IOP Publishing, 2020.

[30] Kumar, M. Keerthi, B. D. Parameshachari, S. Prabu, and Silvia liberata Ullo. "Comparative Analysis to Identify Efficient Technique for Interfacing BCI System." In *IOP Conference Series: Materials Science and Engineering*, vol. 925, no. 1, p. 012062. IOP Publishing, 2020.

[31] L. Tan, N. Shi, K. Yu, M. Aloqaily, Y. Jararweh, "A Blockchain-Empowered Access Control Framework for Smart Devices in Green Internet of Things", ACM Transactions on Internet Technology, vol. 21, no. 3, pp. 1-20, 2021,https://doi.org/10.1145/3433542.

[32] Z. Guo, A. K. Bashir, K. Yu, J. C. Lin, Y. Shen, "Graph Embedding-based Intelligent Industrial Decision for Complex Sewage Treatment Processes", International Journal of Intelligent Systems，2021, doi: 10.1002/int.22540.

[33] Z. Guo, K. Yu, A. Jolfaei, A. K. Bashir, A. O. Almagrabi, and N. Kumar, "A Fuzzy Detection System for Rumors through Explainable Adaptive Learning", IEEE Transactions on Fuzzy Systems, doi: 10.1109/TFUZZ.2021.3052109.

[34] Subramani, Prabu, K. Srinivas, R. Sujatha, and B. D. Parameshachari. "Prediction of muscular paralysis disease based on hybrid feature extraction with machine learning technique for COVID-19 and post-COVID-19 patients." *Personal and Ubiquitous Computing* (2021): 1-14.

[35] Subramani, Prabu, K. Srinivas, R. Sujatha, and B. D. Parameshachari. "Prediction of muscular paralysis disease based on hybrid feature extraction with machine learning technique for COVID-19 and post-COVID-19 patients." *Personal and Ubiquitous Computing* (2021): 1-14.