Classification in Viva Interactions: Distinguishing Questions, Answers, and Statements with BERT Embedding and SVM

Y.S.S.Rohit, V.Hydhava, T.Rushik, Peeta Basa Pati
Department of Computer Science and Engineering, Amrita School of Computing, Bengaluru,
Amrita Vishwa Vidyapeetham, India
*Corresponding author: bp_peeta@blr.amrita.edu, ORCID: 0000-0003-2376-4591
Contributing authors: rohityarabati9@gmail.com, vegesna.hyndhava2005@gmail.com, tangudurushik@gmail.com

***Abstract*—While oral vivas are important in judging student content mastery and presentation approaches, their manual evaluation can be quite difficult due to interfered speech components, intonation differences and the need to differentiate between different types of dialogue. For these challenges, we introduce a model that incorporates machine learning (ML) and natural language processing (NLP) technologies to automatically segment and label viva recordings into the components: question, answer, and statement. The system achieves its best results using SVM with BERT embedding F1-score of 0.81, ensuring accurate segmentation and reliable performance. In exploiting oral and prosodic features, this system improves segmentation precision, reduces transcription time, and enables evaluators to focus on content evaluation. This technology facilitates objective evaluation in ICT(Information communication technology) enfacilitated learning and it is a useful, affordable learning tool for teachers as well as students.**

***Index Terms*—BERT, TF-IDF, SVM, FastText, LIME, Stacking classifier, automated assessment, Online Viva**

I. Introduction

Even for the modern educator and student, oral viva examinations still function as a frontline tool to assess in a realistic way one's understanding and oral presentation skills. Different from written tests, there is an examination that presents a rapid-fire face-to-face questioning and answering session, allowing for the subtle assessment of knowledge. Here the students have to demonstrate not just an overview of the subject but also engage in intellectual discourse when asked a direct question. Spontaneity, articulation, and critical thinking, aspects often lost on the more stringent evaluation of the written test, significantly fall within the forbearance of assessment that this test can add. The viva thus becomes distinctively important in showing a student's potential to relate to complicated issues.

Manual transcription and segmentation wastes valuable resources and introduces bias or errors in the interpretation. For example, even minute errors may misrepresent how the student performed, such as mistaking a question for an answer. Such errors are further amplified by the difference in speech, intonation, and phrasing of speakers. In the case of high-stakes academies, where such interactions contribute to a student's final evaluation, this inconsistency may seem to invite unfair assessment. By exploiting state-of-the-art Machine Learning (ML) and Natural Language Processing (NLP) techniques, an automated system would logically be able to detect questions from answers without human interaction and tremendously save on time and labor resources. Automation looks to promise a highly consistent, uniform system in which assessments can be conducted over viva recordings and automating the transcribing process which allows the evaluators to concentrate more on the content analysis than the time spent doing the transcription. The prospect for such technology is indeed an alluring solution to ease evaluators' burdens and create a fairer error-free assessment process. This would not only help to eliminate inefficiencies and possible biases in the procedural step but also create a standardized assessment method. Accuracy and consistency in using such technology would give equal evaluation to the students while still retaining special benefits of oral examinations: spontaneity and ability to think critically.

The key contributions of this work are

- The original dataset was generated through manual labeling of audio transcript files, thereby enhancing their credibility and usefulness.
- Experiments of different classifiers — Support Vector Machine, Decision Tree, Random Forest, AdaBoost, Naive Bayes, XGBoost, and k-Nearest Neighbors — with different embeddings (TF-IDF, FastText, SBERT, BERT, T5, and RoBERTa) were performed to achieve maximum accuracy.
- A stacking classifier was also put in place to try and enhance classification performance by integrating multiple base classification models.
- Used LIME — an AI interpretable tool — to enhance understanding of model behavior and predictions.

This work contributes towards the United Nations Sustainable Development Goal on Quality Education (UN SDG-4). The rest of the paper is organized as follows: Section II provides a review on existing works in analyzing text classification in ML models. Section III provides the methods to get and use the dataset into models. The results are discussed in Section IV and accompanied with figures of best performing classifiers and embeddings. Lastly, the overall performance and accuracy of the models is summarized in Section V.

II. Related Work

Turning to the tests based on the description, context and different interrogative forms, Kumar et al. [1] underlined that it is necessary to use NLP agents. This method is particularly relevant in structuring viva interactions such that questions and their corresponding answers are well defined.

Bashir et al. [2] proposed that it is necessary to develop efficient techniques using ML and NLP tools for qualifying/evaluating by semantic similarity, coherency, and relevance of free response answers. This can be good to offer an insight into feedback and enhances the efficiency of the separation model. Saraswat et al. [3] provides the idea of the detection of an emotion from the audio files. Knowledge about the feelings of a respondent can be additional when having the background knowledge about the question and answer differentiation.

Vij et al. [4] describe how various sounds or noise in environment impacts, or influences the perceived utterance of words. Therefore, these insights are valuable in improving the possibilities of voice-based question-answer systems. Moondra et al. [5] offers on factors influencing on the reliability of voice recognition systems when exposed to low quality sound environments. This forms and essential component in ensuring that the identified system is capable of functioning as required across the various environments.

Arora et al. [6] provide approaches for how to formulate and how to assess questions in a manner that only good input to the system is accepted. This make questions which have been well framed easily distinguishable from those questions which are not well framed. Rao et al. [7] describe some of the ML algorithms in the removal of unwanted material from text while retaining the content. The efficiency of the expected answer retrieval is also important for the quality of the data which is being processed in a question-answering system.

Feng et al. [8] have also built an adaptive multi-task learning model to propose the integration of the speech and MT tasks altogether in an end-to-end ST model and fix the weights of the tasks at the time of training. To address the problem of modality mismatches, they used optimal transport to align speech and text sequences and thereby obtain more elaborate common embeddings. They demonstrated their experiments in terms of how well this enhanced the translation. Zhang et al. [9], proposed to use Adversarial training for end to end ST in order to bridge the gap of two different modalities of output. In particular, the presented experimental results on the Augmented LibriSpeech English-French and MuST-C GateNMT English-German datasets demonstrated its effectiveness particularly in the low resource setting.

Ali et al. [10] Provide an approach for detecting noise and for reconstructing noisy signals containing randomly-valued impulse noise utilizing stages for analyzing all samples available (input data) in an attempt to compute their similarity to neighbors. They demonstrated PSNR and sound quality performance superior to all but one technique and in particular at lower noise levels. This method proved to have better preservation of the original samples and was characterized by low computational overheads during both training and testing phases, making it almost, on par with, appreciable to implement. Iqbal et al. [11] have studied various signal processing techniques that can improve the speech intelligibility, voice clarity and naturalness in both mobile (cellular) communication as well as internet communication. Regarding their main topic of interest – the methods of speech enhancement digitally, analogously, and in the hybrid manner – the authors indicated strengths and weaknesses of the domain. They only demonstrate how one or another application enhances the quality of speech given clear and powerful inference. Arun et al. [12] proposed the multilingual STT system based on the MFCC and for the classification of MFCCs SVM algorithm has been used. The system utilizes recorded samples of speech and also reference models which, with the aid of mathematical algorithms will be used by the system to develop text with the closest possible resemblance to the input speech. This is quite useful to transcribe speech to text of any word you have in your desired language and it does it nicely. Srinivas et al. [13] proposed new text categorization technique with using multitype features coselection for clustering (MFCC) based on mining association rule for generating text classifiers. Because the dimensionality of text high can be disadvantageous in categorization MFCC is applied as a technique in feature reduction. The method delivers high quality document vectors and enhances the results of clustering. The testing indicated that it has higher or similar accuracy and F1 scores than Decision Tree classifiers despite the higher number of features. The cited works employ different strategies to improve the speech recognition systems and question answering systems. They explain the applications of NLP and machine learning for making interactions, assessing the answers' content based on semantic value, and considering issues like background noise, emotion sensing, and question asking. Ways on how to enhance intelligibility of speech, maintain sound quality, and minimize the effects of background noise on speech are also discussed. In addition, techniques for solving the problems of the modality gaps in the speech-to-text tasks and approaches for utilizing multilingual systems for bettering the transcription of the text have been stressed. However, even with these new technologies, most studies do not address the application of explainable AI models and their impact on the interpretability and trustworthiness of complex systems built with machine learning. In this paper, we tackle this problem by demonstrating how to train speech classifiers and interpret the results with protective methods like LIME. This increases the correlation between the interpretation of the results and the actual problem being solved.

III. Methodology

Fig. 1 illustrates the workflow for viva classification, starting with audio data collection, noise reduction, segmentation, and tokenization. Embedded techniques, such as BERT or FastText, are used for effective speech representation. Machine learning models like SVM, Random Forest, and XGBoost evaluate classified speech segments, with a stacking classifier improving performance. LIME is integrated for model explainability, leading to final predictions.

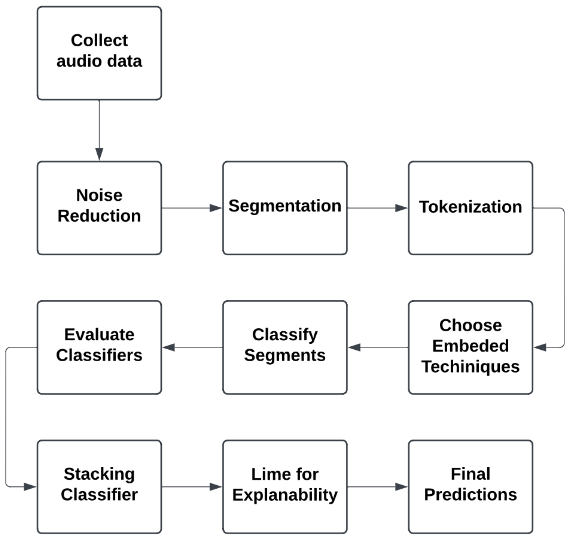

Fig. 1: Workflow for Speech Classification Using Embedding Techniques

A. Data Collection and Preprocessing

The process of developing the system starts with capturing viva recordings, each viva recording is about 12 minutes and has around 3 speakers in it, this is the actual raw audio data that is collected. The noise reduction step is therefore conducted with a view of enhancing the clarity of the given speech, thus reducing the level of background noise. This makes sure that the subsequent processes work on clear audio and hence produce the correct results. Then that data is converted into text using Transcriber in Microsoft Word. Then comes the process of segmentation where the text is divided into segments in terms of questions, answers, or statements. The text data has 1725 statements, 803 questions and 1009 answers.

B. Tokenization and Embedding Techniques

After decomposing the data into parts and labeling, different types of embedding techniques were used to find which one gives the highest accuracy.

- **BPE (Byte Pair Encoding):** BPE breaks words into sub-word units to handle (OOV) words well. This functionality is particularly beneficial in speech-related tasks, where lexical variation and infrequent words are highly present. The Bisht and Gupta-researched et al. [14] emphasized the importance of subword tokenization to overcome (OOV) challenges that further enhance translation accuracy. This adaptability to unfamiliar words makes BPE highly suitable for speech data wherein different patterns of lexical variations must be represented economically.
- **BERT(Bidirectional Encoder Representations from Transformers):** BERT has had to employ bidirectional attention in producing context-aware token embeddings for its exceptional performance in undertaking tasks that require deep contextual understanding. According to Kumar et al. [15], BERT is quite effective with complex language patterns and recognizing subtle relationships within text. Its advanced contextual processing capabilities make it the perfect tool for analyzing sophisticated speech data, that might easily escape conventional methods.
- **FastText:** FastText words are represented with character n-grams, which capture effective morphological patterns. That's why it is useful for small data also and in scenarios where computational resources are constrained. The work [15] stated that lightweight embedding models, such as FastText, are suitable for resource-constrained environments for news applications." Its lightweight architecture, coupled with its aptness to work with smaller datasets makes FastText a sensible choice for resource-constrained speech analysis tasks.
- **T5(Text-to-Text Transfer Transformer):** T5 is a transformer-based text-to-text model that reformulates NLP tasks as such, thus allowing for a unified approach toward all and sundry applications. It works well with generating those embeddings where the semantic and syntactic information entailed can be very useful in several applications, such as in speech analysis and classification. According to Sandhiya et al. [16], transformer-based embeddings work wonderfully in retaining close relationships within text data. In that respect, because T5 can process tasks in a single framework, it is very efficient for applications that require contextual understanding and transformation, especially transcribing and analysis of speech data.
- **SBERT:** SBERT is an extension of BERT and is fine-tuned for the generation of sentence embeddings that capture semantic relationships between sentences. It further boosts the classification of speech segments by contextualizing the relationships between spoken phrases. Sandhiya et al. [16] demonstrated the importance of embeddings in capturing semantic meaning to derive better classification results. SBERT's capacity to understand semantic relationships within contextually dependent sentences makes it indispensable for tasks requiring nuanced understanding of spoken content.
- **RoBERTa (A Robustly Optimized BERT):** RoBERTa indicates the advanced BERT model that concentrates on training strategies such as longer sequence and larger datasets processing. This also supports robust analysis for complex speech data handling. Nair et al. [17] shown the optimized embedding technique such as RoBERTa performs better than simple models in classification. Its capacity to process with fine details for huge datasets is very crucial for the analysis of speech segments.
- **TF-IDF (Term Frequency-Inverse Document Frequency):** A statistical method that measures the im-

portance of words in a document compared to their occurrence in a corpus. This technique has widely been used in text classification because it's simple and requires minimal computation power. Nair et al. [17] highlights its effectiveness in structured text processing applications that rely on key word extraction. Its low computational overhead and easy implementation make it a suitable candidate for applications requiring efficient and scalable keyword extraction.

C. Classification and Model Training

The extracted embeddings are then passed through seven different classifiers to determine which model gives the highest performance in segmenting speech as a question or answer or a statement. The classifiers used include Support Vector Machine (SVM), Decision Tree, Random Forest, AdaBoost, Naive Bayes, XGBoost, and k-Nearest Neighbors (kNN). Each of the classifiers is tuned and trained on a labeled dataset, and the optimal parameters were very carefully calibrated to give the best performance for each of the models.

D. Classifiers and Best Hyperparameters

Table I provides an overview of the best hyperparameters and embedding techniques used for different machine learning models in the speech classification task. For instance, models like SVM, Random Forest, and XGBoost relied on BERT embeddings to capture context more effectively, while Decision Tree and Naive Bayes made use of TF-IDF for feature extraction. Meanwhile, kNN utilized FastText embeddings to focus on subword-level details. This table emphasizes how careful tuning of parameters and the choice of advanced embedding techniques contribute to improving the performance of these models.

The last step before classification would be a quantitative interclass separable representation, which would allow a judge as to whether the projection features distinctly separate them into groups. This includes computation of means for each class and considering the density imbalance between the variables that might affect the differentiation of classes properly. Minkowski distance parameters are also fine-tuned to improve distance classification performance of models such as kNN.

Each model was evaluated with metrics for precision, recall, and F1-scores obtained from confusion matrices. Thus, it allows assessment and comparison of the efficiency of each classifier working on cases within a speech category. For instance, kNN classifier performance was tested with various values of k, set out to find an ideal trade-off between bias and variance, based on the smallest amount of classification errors.

Once classification is done, the stacking classifier works next and the cross-validation prediction is used to train a meta-model thus improving the initial prediction. This step translates as an effort to make sure that the meta-model equips the organisation with the combined knowledge of the base models, achieve through using predictions computed from data that are unknown to the base models. The stacking process ensures that patterns that were otherwise unnoticed are identified and this makes the result set more accurate. This refinement allows the meta-model to generate the last predictions containing the features which all base models performed well to achieve higher accuracy levels and to accommodate difficult cases much better.

TABLE I: Best Parameters and Embeddings for Different Models

Model	Best Parameters	Embedding
SVM	`{'kernel': 'poly', 'gamma': 0.01, 'degree': 4, 'coef0': 0.5, 'C': 0.46415888336127775}`	BERT
Decision Tree	`{'splitter': 'random', 'min_samples_split': 10, 'min_samples_leaf': 1, 'max_features': None, 'max_depth': 20, 'criterion': 'entropy'}`	TF-IDF
Random Forest	`{'n_estimators': 200, 'min_samples_split': 10, 'min_samples_leaf': 1, 'max_features': 'sqrt', 'max_depth': 40, 'bootstrap': True}`	BERT
AdaBoost	`{'n_estimators': 100, 'learning_rate': 0.01, 'estimator__min_samples_split': 2, 'estimator__min_samples_leaf': 4, 'estimator__max_depth': 3}`	BERT
Naive Bayes	`{'fit_prior': True, 'alpha': 0.5}`	TF-IDF
XGBoost	`{'subsample': 0.8, 'reg_lambda': 0.01, 'reg_alpha': 0.01, 'n_estimators': 300, 'max_depth': 6, 'learning_rate': 0.05, 'gamma': 0, 'colsample_bytree': 1.0}`	BERT
kNN	`{'weights': 'distance', 'n_neighbors': 8, 'metric': 'euclidean'}`	FastText

Finally, we apply the LIME method (Local Interpretable Model-agnostic Explanations) which renders the analysis comprehensible too. Such methods make it possible to analyze in detail the output of each model and present for perception serpents that were guessed wrong. LIME also serves to identify the major determinants, and the application of these determinants allows us to reach final predictions with all necessary corrections. The findings of this study demonstrate the most successful model as well as the model features employed in the classification process.

IV. Results and Discussion

For the purpose of comparing tokenization processes, different tokenization processes have been applied such as TF-IDF, Byte Pair-Encoding (BPE), SBERT, BERT, RoBERTa, FastText, and T5 in terms of classification using several models like Support Vector Machine (SVM), Decision Tree, Random Forest, AdaBoost, Naive Bayes, XGBoost, and k-Nearest Neighbors (kNN). Every tokenization process will reproduce embeddings differently, due to which different classification models will reach different levels of classification accuracy when classifying a speech segment in one of the three classes: Question, Answer, and Statement. BERT, TF-IDF and T5 consistently demonstrated the highest accuracy in classification across classification models, mainly using SVM and Random Forest. This statistical approach revealed the significant importance of representational learning models in contrast with traditional statistical representations.

The trends and figures have allowed us to formulate the conclusion that the embedding techniques and the classifiers behave differently during the speech classification tasks. While the comparison ofF1-scores of Fig. 2 (TF-IDF, BERT, and T5 embeddings method with SVM) illustrates the BERT and T5 models to be more effective than other models, this is only primarily because they are able to utilize context better when more information is provided during training as well as during testing. A similar trend can be noted in Fig. 3, IT is a comparison of F1-scores of BERT embeddings across the classifiers in this study, where it is noted that Random Forest and SVM still remain at the top for several logits because they are more capable of the complex features embedded in each of the tokenized speech components. Such patterns highlight the general concept that speech data, just like in any other classification problem, requires attention to the choice of embeddings and the classifiers.

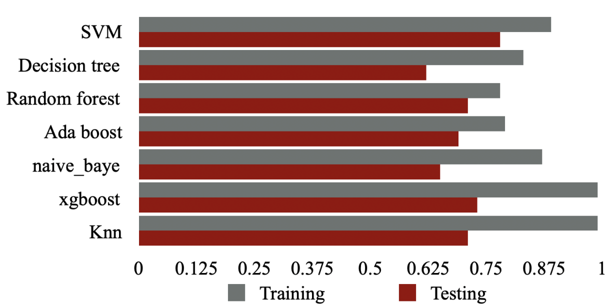

Fig. 2: Performance comparison for various classifiers with BERT embedding using F1 score

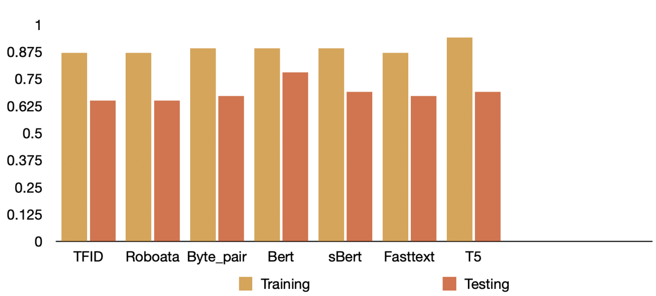

Fig. 3: Performance comparison for various embeddings with SVM classifier using F1-score

Local Interpretable Model-Agnostic Explanations (LIME) offered useful perspectives regarding classification workings of the model which was aided by the significance of words in determining speech segments. As shown in Fig. 4(a), more important words like "What" and "is" have the most weight and serve the purpose of marking questions. However, "speech" and "processing" are still meaningful words for classification but are of lesser weight. In Fig. 4(b), it appeared that the classifying probabilities were more predominant towards "Answer" (0.68) and "Statement" (0.27) leaving a weak 0.04 for "Question," which implied that the segments were hard to identify as when, for instance, the terms "outcome" and "leakage" were not very helpful indicators. This points out the problem with the model: sometimes it has problems with contextual understanding.

Exactly the same kind of situation was also observed in Fig. 4(c) when the model had no problems classifying the statements as such because numerous discourse markers and words such as "Yeah" and "You are saying" occupied high ranks in the classification. Here the most prevalent probability was given to "Statement" (0.52), whereas "Answer" and "Question" probabilities were placed at slightly lower levels of 0.42 and 0.05 respectively. As these results are shown in Fig. 4(c), the model depends on some specific discourse markers to differentiate types of speech, and this shows that there is a place for improvement as well, exactly the opposite.

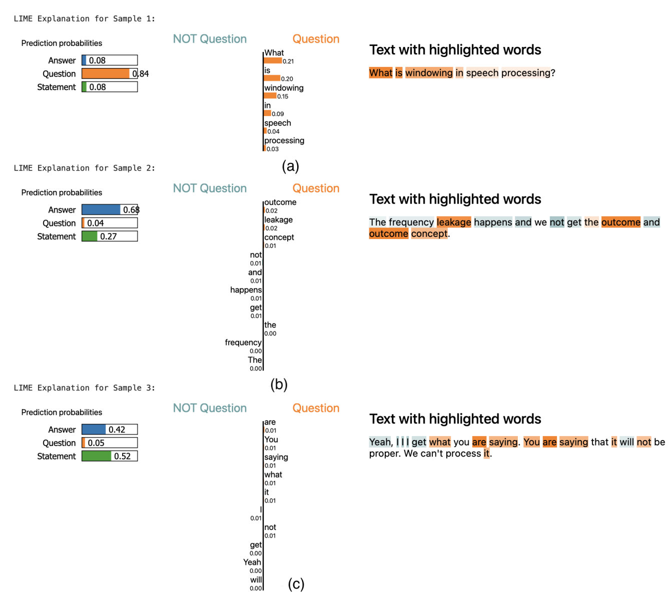

Fig. 4: Word influence analysis for classification on sample data using LIME. (a) Question sample, (b) Answer sample, (c) Statement sample

Another explanation was obtained with the assistance of the BERT similarity model that enabled text classification to be more informed due to raised context level. The Local Interpretable Model-agnostic Explanations (LIME) technique also enhanced interpretability of model predictions with LIME capturing important aspects in each of the segments that brought about high confidence in the classification. By employing this interpretability step, the understanding and optimization of the performance of the model would be effortless since the variables that mattered most in making the predictions were highlighted.

In addition to these findings, Stacking ensemble model was tested for the classification by obtaining the F1-score of 0.80 over the data set, making it reliable in terms of model performance. This is because the individually tuned RandomForest, SVM, KNeighborsClassifier, and XGBClassifier each were able to find local decision regions and when combined with the GradientBoostingClassifier, were able to identify all the high-level patterns that exist in the data. Extending the search for the hyperparameter and the cross-validation added to the stable and reliable performance of this diverse stacking.

The last phase of this research, where precision, recall, and F1-score metrics have been employed, demonstrates the effectiveness of using tokenization and embedding techniques of speech rather further in this scope. Complex models such as BERT and some tokenization methods improved the performance of the system to some extent in differentiating between the three segments of speech i.e. questions, answers, and statements. Also, use of contextual knowledge and post-processing mechanisms facilitated the elimination of the ambiguities and resulted in better and cleaner predictions.

V. Conclusion

This work demonstrates that transformer-based embeddings, BERT and sBERT, perform better compared to traditional approaches such as TF-IDF and FastText. Besides, when combined with SVM classifier, these fine-tune their performances significantly. SVM is shown to be the more reliable classifier in most cases and scenarios using embeddings. Ensemble methods such as Random Forest and XGBoost are very good alternatives if a complex dataset exists. The results from LIME highlighted the importance of specific vocabulary and discourse features in the classification task, which may thus be improved to deal better with ambiguous situations. To sum up, this study highlights the need for balancing efficiency, accuracy, and interpretability in speech classification efforts by tailor-mending methods to specific data and resource constraints.

Future work can focus on improving contextual handling of ambiguous speech segments by integrating dynamic embeddings and real-time processing capabilities. Lightweight transformer models or hybrid approaches can be explored to balance efficiency and semantic depth. Additionally, expanding these systems for multilingual and domain-specific datasets can enhance their scalability and applicability to diverse real-world scenarios.

Acknowledgment

Authors would like to express sincere gratitude to Amrita Vishwa Vidyapeetham for the resources and supportin carrying out this research. Authors also acknowledge the usage of ChatGPT for editing assistance and to Google Colab for experimentation.

References

[1] A. Kumar, D. Singh, A. Kharadi and M. Kumari, "Automation of Question-Answer Generation," 2021 Fourth International Conference on Computational Intelligence and Communication Technologies (CCICT), Sonepat, India, 2021, pp. 175-180, doi: 10.1109/CCICT53244.2021.00043.

[2] M. F. Bashir, H. Arshad, A. R. Javed, N. Kryvinska and S. S. Band, "Subjective Answers Evaluation Using Machine Learning and Natural Language Processing," in IEEE Access, vol. 9, pp. 158972-158983, 2021, doi: 10.1109/ACCESS.2021.3130902.

[3] S. Saraswat, S. Bhardwaj, S. Vashistha and R. Kumar, "Sentiment Analysis of Audio Files Using Machine Learning and Textual Classification of Audio Data," 2023 6th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2023, pp. 1-5, doi: 10.1109/ISCON57294.2023.10112195.

[4] D. Vij, Y. Yogesh, D. Srivastava and H. Shankar, "Detection of Acoustic Scenes and Events using Audio Analysis – A Survey," 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 316-320, doi: 10.1109/ICACITE57410.2023.10183195.

[5] A. Moondra and P. Chahal, "Voice Feature Extraction Method Analysis for Speaker Recognition with Degraded Human Voice," 2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2023, pp. 385-388, doi: 10.1109/ICAC3N60023.2023.10541716.

[6] U. Arora, N. Goyal, A. Goel, N. Sachdeva and P. Kumaraguru, "Ask It Right! Identifying Low-Quality questions on Community Question Answering Services," 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 2022, pp. 1-8, doi: 10.1109/IJCNN55064.2022.9892454.

[7] M. C. Rao, K. C. Yelavarti and N. P. Kalyan, "A Framework for Hate Speech Detection using Different ML Algorithms," 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2023, pp. 960-967, doi: 10.1109/ICOEI56765.2023.10125942.

[8] Feng, X., Zhao, Y., Zong, W. Xiaona Xu, Adaptive multi-task learning for speech to text translation. J AUDIO SPEECH MUSIC PROC. 2024, 36 (2024). https://doi.org/10.1186/s13636-024-00359-1

[9] Hao Zhang, Xukui Yang, Dan Qu, Zhen Li, Bridging the cross-modal gap using adversarial training for speech-to-text translation, Digital Signal Processing, Volume 131, 2022, 103764, ISSN 1051-2004,

[10] Ali Awad, Impulse noise reduction in audio signal through multi-stage technique, Engineering Science and Technology, an International Journal, Volume 22, Issue 2, 2019, Pages 629-636, ISSN 2215-0986,

[11] J. L. M. Iqbal and A. Karthik, "Speech Enhancement Techniques in the Digital, Analog, and Hybrid Domains," 2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2022, pp. 1-3, doi: 10.1109/ICCCI54379.2022.9740883.

[12] Dubey, A.K., Arora, Y., Gupta, N., Yadav, S., Jain, A., Verma, D. (2024). Multi-featured Speech Emotion Recognition Using Extended Convolutional Neural Network. IACC 2023. Communications in Computer and Information Science, vol 2053. Springer

[13] M. Srinivas, K. P. Spreethi, E. V. Prasad and S. Anitha Kumari, "MFCC and ARM algorithms for text categorization," 2008 International Conference on Computing, Communication and Networking, Karur, India, 2008, pp. 1-6, doi: 10.1109/ICCCNET.2008.4787780.

[14] A. R. Nair, S. M and D. Gupta, "Comparative Analysis of Word Embeddings for Text Classification in Spark NLP," 2023 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), Mysuru, India, 2023, pp. 130-136, doi: 10.1109/CCEM60455.2023.00027.

[15] Prasanna Kumar R, Bharathi Mohan G, Elakkiya R and Varsha P, "A Comparison of Word Embeddings for Comment Toxicity Detection: Detection Power of Computer," 2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI), Greater Noida, India, 2023, pp. 393-398, doi: 10.1109/ICCSAI59793.2023.10421356.

[16] Sandhiya B and S. Santhanalakshmi, "News Article Topic Classification Using Embeddings," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-7, doi: 10.1109/ICCCNT56998.2023.10306603.

[17] A. Bisht, D. Gupta and S. Parida, "Guided Transformer for Machine Translation: English to Hindi," 2023 IEEE 20th India Council International Conference (INDICON), Hyderabad, India, 2023, pp. 636-641, doi: 10.1109/INDICON59947.2023.10440876.