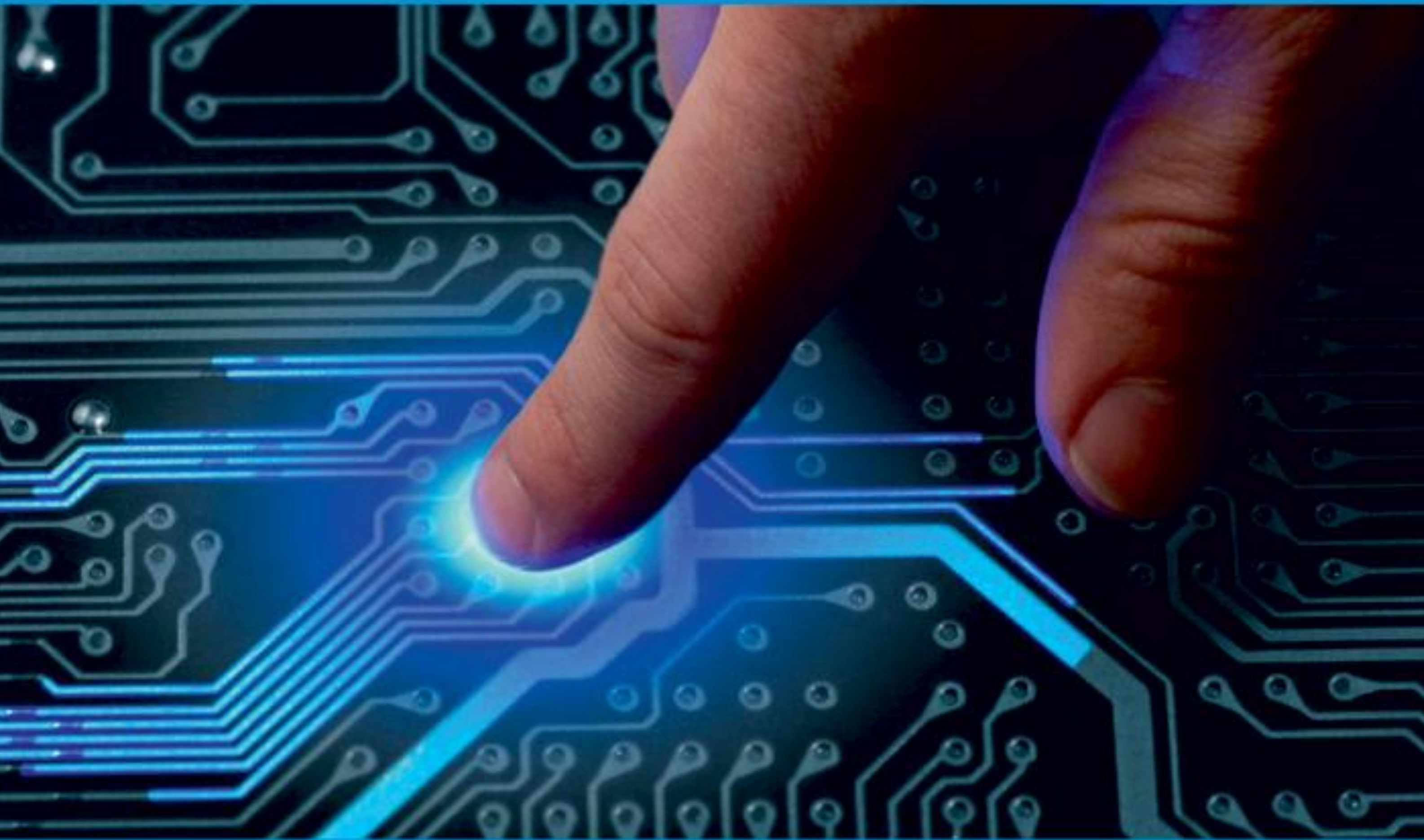
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205256 |

CNN based Intelligent Question Answer Generating System for Visually Impaired Student

Dr. (Mrs.) Archana P. Kale, Abhishek Mandave, Pratik Walunj, Shubham Chikane, Sakhib Shaikh

Professor, Department of Computer., MES Wadia College of Engineering Pune, India

UG Student, Department of Computer MES Wadia College of Engineering Pune, India

ABSTRACT: The rapid evolution of the Internet has revolutionized the landscape of teaching and learning, ushering in a digital era that strives to make knowledge universally accessible. In this context, it is imperative to address the needs of visually impaired students, ensuring they have equal access to the wealthof information available on the internet. To facilitate this, our proposed methodology leverages the synergies between Text-to-Text-Transfer-Transformer contextual understanding and Convolutional Neural Networks (CNN) for semantic comprehension.

Additionally, it integrate textual image recognition to enable efficient text extraction and question-answer generation. The core innovation lies in empowering voice assistants to dynamically generate contextually relevant and subjective questions tailored to individual user preferences. This approach not only enhances the learning experience for visually impaired students but also extends to the broader domain of personalized voice assistant interactions. To ensure informative and human-like responses, we incorporate a robust text-to-speech system that synthesizes answers using T5.

KEYWORDS: Subjective question generation, Transformers, T5, CNN, Text-to-speech, Natural language processing, Voice assistant.

I. INTRODUCTION

In the pursuit of fostering an inclusive and equitable educational environment, the recognition of education as a fundamental right stands as a cornerstone. However, the journey towards universal access to quality learning resources becomes intricate when considering the diverse needs of individuals, particularly those with physical disabilities. Traditional text based learning materials, while invaluable, often present substantial barriers to inclusivity. The inherent limitations of these resources can impede the educational progress of individuals with disabilities, hindering their ability to fully engage with and benefit from the learning process. The technological landscape, however, offers a promising avenue for overcoming these challenges and ensuring that educational opportunities are accessible to all. This report delves into the integration of cutting-edge technologies such as Convolutional Neural Networks (CNN), Bidirectional Encoder Representations from Transformers (T5), Natural Language Processing (NLP), Textto-Speech (TTS), and Machine Learning (ML) as catalysts for fostering inclusivity in education. By exploring the synergies between these technologies and educational accessibility, we aim to shed light on the transformative potential they hold in dismantling barriers and enhancing the learning experience for individuals with physical disabilities. Through an in-depth examination of the application of these technologies, we endeavour to contribute insights that can inform the development of inclusive educational practices, ensuring that no one is left behind on the path to knowledge and empowerment.

II. RELATED WORK

In the realm of text extraction and question generation, machine learning techniques play a pivotal role, facilitating the automatic extraction of meaningful information from images. These techniques leverage convolutional neural networks (CNNs) for feature extraction and recurrent neural networks (RNNs) for decoding extracted features into coherent text. Machine learning approaches widely provide beneficial approach for key phrase extraction used pre trained transformers like T5.T5 model architecture with multi-layer transformer with self-attention layer applied on datasets like NEP dataset.

A CNN based text detection using dataset containing English text, CNN model detects several text blocks by applying edge descriptors in different blocks CNN is pre-trained by convolutional sparse auto encoder (CSAE). Convolutional layer in CNN model extracts useful features from textual images for detecting the text and the language. The extracted text conversion into audio output to easily listen to the generated content.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205256 |

Numerous pre-trained Text-to-Speech (TTS) engines are accessible for converting text into spoken audio. Widely used Python library for TTS is Google's Text-to-Speech API, which can be accessed through the gTTS library.

III. PROPOSED ALGORITHM

The proposed methodology of project involves user interaction and text detection in natural scene, extracted text is further preprocessed by transformer for question answer generation text to speech model provides a better way to dictate every question and answer to the visually impaired students. Methodology can be described in four phases which involves text detection and extraction from the textual images, text pre-processing, and question answer generation from the extracted context from the textual image and at last text to speech conversion. The general architecture of the proposed intelligent question answer generator is illustrated in upcoming figures. The framework is based on three components:

1. The CNN component feature extraction from images.
2. LSTM for sequential text processing,
3. Question Answer generator using T5,
4. Text to Speech system for audio output of generated question answers.

A. *CNN LSTM CTC Layer:*
The CNN LSTM CTC layer which is a common architecture used in field of deep learning for sequence-to-sequence tasks in the context of speech recognition or optical character recognition (OCR). It is generally used in image classification and text retrieval process. CNN extracts the features from the textual images, this process is similar to Optical Character recognition.

1. Text Detection Fig. 2. Recognition (OCR) which is also used in text extraction from images.
1) Convolutional Neural Network (CNN): CNNs are typically used for extracting features from input data, generally used in image processing tasks. In the context of sequence to sequence tasks, CNNs can be used to process sequential data such as spectrograms in speech recognition or text from images similar to OCR, to capture relevant patterns.

2) Long Short-Term Memory (LSTM): LSTMs are a type of recurrent neural network (RNN), LSTMs are designed to capture long-term dependencies in sequential data. LSTMs are particularly useful when dealing with sequences of textual data if variable length and have been widely used in natural language processing (NLP) and speech recognition like tasks.

3) Connectionist Temporal Classification (CTC) Layer: CTC is a loss function that is used in sequence-to-sequence tasks where the alignment between the input and output feature sequences is not known. It is particularly applied in tasks like speech recognition or handwritten text recognition, where the length of the input sequence might not align perfectly with the length of the target sequence. When this components are combined together it will form architecture which will extract the feature from input data, where LSTM capture the sequence to sequence dependencies, and CTC layer align the predicted layer with original sequence.

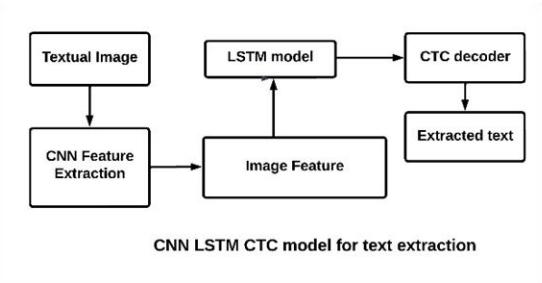


Figure 1.Text Extraction

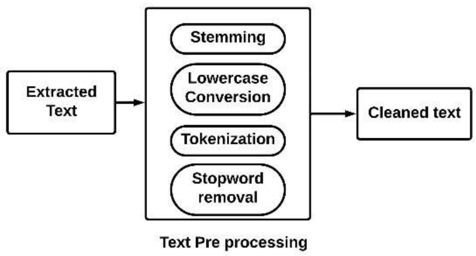


Figure 2. Text Pre-processing

B. *Text Pre-processing:-*
Text pre-processing is a important step in natural language processing (NLP) and machine learning tasks involving textual data. The extracted text from the textual images need to be cleaned before sending as input for question answer

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

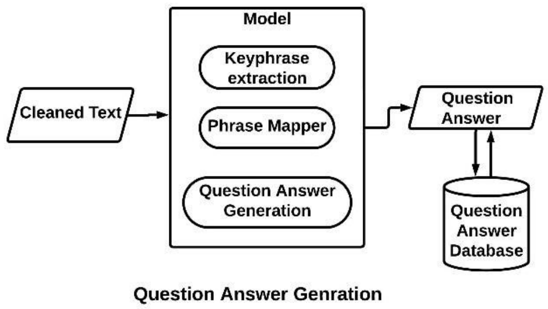
| DOI: 10.15680/IJIRCCE.2024.1205256 |

generation. The most important work of text pre-processing is to clean, normalize, and transform raw text data into a format that is suitable for model to generate question and answer. Text pre-processing involves stemming and lemmatization, special character removal, stop word removal, Tokenization. These pre-processing steps helps in cleaning and transforming raw text data into a format that is more suitable for NLP tasks, such as text analysis, machine learning model training.

C. Question Answer Generation:-

Question answering (QA) generation is a task in natural language processing (NLP) that involves a model that can help to automatically generate questions based on a given context or passage.

1) Transformer based language models: T5-base is one of the pre-trained transformer based language model developed by Google. Unlike other traditional language models that process text in a left to right or right to left manner,t5 is designed to understand the context in both the directions (bidirectional). It considers both the left and right context for each word it allows model to capture important contextual information. BERT is one of the pre-trained language model trained on a massive amount of textual data using unsupervised learning. During this pre-training phase model learns to predict missing words in a sentence by considering the context on both sides of the gap. After pre-training phase the BERT can be fine-tuned for specific tasks on new dataset which contains question answering. Fine-tuning involves training the model on a labelled dataset where the model learns to generate accurate answers to questions generated based on input passages.



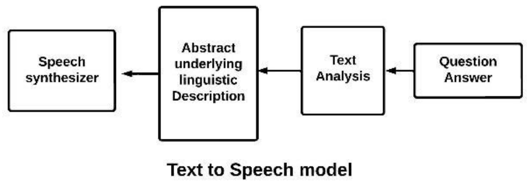


Figure 3.Question Answer Generation

Figure 4.Text to Speech Model

D. Text to Speech engine:-

The extracted text is need to be converted in audio output for visually impaired students so that they can easily listen the generated content. There are several pre-trained Text to Speech (TTS) engines available that can be use to convert text into spoken audio. One of the popular library for TTS in Python is Google's Text-to-Speech API it is available through the gTTS library. This library converts text to speech using Google's pre-trained models.

IV. SIMULATION RESULTS

To demonstrate the working of automated question answer generating system as per proposed system, the input textual, content goes through different layers and these layers generates feature sequences which are further used for question answer generation using T5 and BERT based models for important text extraction and summarization purpose. The summary extracted from the input context is passed through T5 model which will generate question by focusing on important key points which has been summarized by BERT based summarizer pipeline. These pipeline consists of different models such as BERT summarizer,t5-base,t5-large,BART models like facebook bart-large cnn, facebook bart-large and google pegasus-xsum also outperforms BERT summarizer model.

The network showed in Figure 1. is able to transmit 22 packets if total transmission energy metric is used and 17 packets if used maximum number of hops metric. And the network lifetime is also more for total transmission energy. It clearly shows in Figure 2. that the metric total transmission energy consumes less energy than maximum number of hops. As the network is MANET means nodes are mobile and they change their locations. After nodes have changed their location the new topology is shown in Figure 3. and energy consumption of each node is shown in Figure 4. Our

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205256 |

results shows that the metric total transmission energy performs better than the maximum number of hops in terms of network lifetime, energy consumption and total number of packets transmitted through the network.

Choosing a summarizer which will perform well on the contextual input by considering constraints like length of contextual input, model size and its performance. Evaluation metrics like BLEU (Bilingual Evaluation Understudy), ROUGE (Recall-Oriented Understudy for Gisting Evaluation) used for calculating the performance metrics of generated summary on reference summary. BLEU metrics is based on n grams, BLEU calculates the percentage of n gram (sequence of words) in the generated summary matching n grams in the reference summary. Scores between 0.4 to 0.6 are acceptable in summarization task while score above 0.7 can be considered as good performance. ROGUE model calculates overlap between generated summary and reference summary, it measures the recall of the generated summary which measures how much information from the generated summary present in reference summary. There are different ROGUE variants such as ROGUE 1,ROGUE 2 and ROUGE-L.ROGUE 1 measures the performance of summarizer by overlapping of single word (unigrams) between the generated and referenced summary.ROGUE-2 works like ROGUE-1 but instead of unigrams it uses bigrams overlap that is overlap of sequence of consecutive two words, higher ROGUE-2 score indicates better overlap of consecutive words, ROUGE-L measures the longest common subsequence (LCS) between n grams of generated and reference summary .After executing ROGUE and BLEU evaluation metrics on generated summaries using summarizer models like t5- base, BART,BERT summarizer following results are obtained.

2*Metric	Scores		
	Recall (r)	Precision (p)	F1 Score (f)
ROUGE-1	0.8305	0.98	0.8991
ROUGE-2	0.7632	0.9206	0.8345
ROUGE-L	0.8305	0.98	0.8991
BLEU	-	-	0.7956

TABLE I
ROUGE AND BLEU SCORES FOR BART-LARGE-CNN

2*Metric	Scores		
	Recall (r)	Precision (p)	F1 Score (f)
ROUGE-1	0.1525	0.4737	0.2308
ROUGE-2	0.0263	0.0909	0.0408
ROUGE-L	0.1186	0.3684	0.1795
BLEU	-	-	0.3442

TABLE III
ROUGE AND BLEU SCORES FOR GOOGLE PEGASUS-XSUM

2*Metric	Scores		
	Recall (r)	Precision (p)	F1 Score (f)
ROUGE-1	0.6102	0.8571	0.7129
ROUGE-2	0.5	0.717	0.5891
ROUGE-L	0.6102	0.8571	0.7129
BLEU	-	-	0.6083

TABLE II
ROUGE AND BLEU SCORES FOR T5-BASE

adjustbox 2*Metric	Scores		
	Recall (r)	Precision (p)	F1 Score (f)
ROUGE-1	0.9048	0.6333	0.7451
ROUGE-2	0.7500	0.5172	0.6122
ROUGE-L	0.9048	0.6333	0.7451
BLEU	-	-	0.3894

TABLE IV
ROUGE AND BLEU SCORES FOR BERT SUMMARIZER

V. CONCLUSION AND FUTURE WORK

In this paper we have discussed comprehensive solution for improving the educational experience for visually impaired students. Intelligent question answer generator includes setting edge technologies like CNN, RNN and Text to text transfer transformer (T5).CNN model used for extracting text from the images, RNN used for sequence to sequence text handling and for understanding temporal context of textual image. CTC model helps in sequence to sequence task in feature extraction from the image. T5 model performed pivotal role in generating high quality of questions and answers.T5's summarizer attention mask played important role in important sentence extraction and generating summary of the long summary which is further provided to T5-base model which is pre trained on SQUAD dataset. The inclusion of text to speech makes it easier to read aloud generated questions and answers for visually impaired students, this breakdowns barrier between educational content and visual disability, making it easier for students to learn and self-evaluation.

REFERENCES

1. Intelligent Question Answering in Restricted Domains Using Deep Learning and Question Pair Matching LIN-QIN CAI , (Member, IEEE), MIN WEI , SI-TONG ZHOU , AND XUN YAN.
2. M. A. Kia, A. Garifullina, M. Kern, J. Chamberlain and S. Jameel, "Adaptable Closed-Domain Question Answering Using Contextualized CNN-Attention Models and Question Expansion," in IEEE Access, vol. 10, pp. 45080-45092, 2022, doi: 10.1109/ACCESS.2022.3170466.
3. L. H. Son, A. Kumar, S. R. Sangwan, A. Arora, A. Nayyar and M. AbdelBasset, "Sarcasm Detection Using Soft Attention-Based Bidirectional Long Short-Term Memory Model With Convolution Network," in IEEE Access, vol. 7, pp. 23319-23328, 2019, doi: 10.1109/ACCESS.2019.2899260.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205256 |

4. D. R. CH and S. K. Saha, ”Automatic Multiple Choice Question Generation From Text: A Survey,” in IEEE Transactions on Learning Technologies, vol. 13, no. 1, pp. 14-25, 1 Jan.-March 2020, doi: 10.1109/TLT.2018.2889100.
5. T. Shao, Y. Guo, H. Chen and Z. Hao, ”Transformer-Based Neural Network for Answer Selection in Question Answering,” in IEEE Access, vol. 7, pp. 26146-26156, 2019, doi: 10.1109/ACCESS.2019.2900753.
6.] W. T. Alshammari and S. AlHumoud, ”TAQS: An Arabic Question Similarity System Using Transfer Learning of BERT With BiLSTM,” in IEEE Access, vol. 10, pp. 91509-91523, 2022, doi: 10.1109/ACCESS.2022.3198955.
7. N. Alsaaran and M. Alrabiah, ”Classical Arabic Named Entity Recognition Using Variant Deep Neural Network Architectures and BERT,” in IEEE Access, vol. 9, pp. 91537-91547, 2021, doi: 10.1109/ACCESS.2021.3092261.
8. M. A. Khan, P. Paul, M. Rashid, M. Hossain and M. A. R. Ahad, ”An AI-Based Visual Aid With Integrated Reading Assistant for the Completely Blind,” in IEEE Transactions on Human-Machine Systems, vol. 50, no. 6, pp. 507-517, Dec. 2020, doi: 10.1109/THMS.2020.3027534.
9. R. Devika, S. Vairavasundaram, C. S. J. Mahenthar, V. Varadarajan and K. Kotecha, ”A Deep Learning Model Based on BERT and Sentence Transformer for Semantic Keyphrase Extraction on Big Social Data,” in IEEE Access, vol. 9, pp. 165252-165261, 2021, doi: 10.1109/ACCESS.2021.3133651.
10. S. Zhang, H. Tan, L. Chen and B. Lv, ”Enhanced Text Matching Based on Semantic Transformation,” in IEEE Access, vol. 8, pp. 30897-30904, 2020, doi: 10.1109/ACCESS.2020.2973206.
11. Y. Zhang, S. Nie, S. Liang and W. Liu, ”Robust Text Image Recognition via Adversarial Sequence-to-Sequence Domain Adaptation,” in IEEE Transactions on Image Processing, vol. 30, pp. 3922-3933, 2021, doi: 10.1109/TIP.2021.3066903.
12. X. Mu and A. Xu, ”A Character-Level BiLSTM-CRF Model With Multi-Representations for Chinese Event Detection,” in IEEE Access, vol. 7, pp. 146524-146532, 2019, doi: 10.1109/ACCESS.2019.2943721.
13. C. -W. Tseng, J. -J. Chou and Y. -C. Tsai, ”Text Mining Analysis of Teaching Evaluation Questionnaires for the Selection of Outstanding Teaching Faculty Members,” in IEEE Access, vol. 6, pp. 72870-72879, 2018, doi: 10.1109/ACCESS.2018.2878478.
14. W. Yu, M. Yi, X. Huang, X. Yi and Q. Yuan, ”Make It Directly: Event Extraction Based on Tree-LSTM and Bi-GRU,” in IEEE Access, vol. 8, pp. 14344-14354, 2020, doi: 10.1109/ACCESS.2020.2965964.
15. G. Xu, Y. Meng, X. Zhou, Z. Yu, X. Wu and L. Zhang,”Chinese Event Detection Based on Multi-Feature Fusion and BiLSTM,” in IEEE Access, vol. 7, pp. 134992-135004, 2019, doi: 10.1109/ACCESS.2019.2941653.
16. X. Ren, Y. Zhou, J. He, K. Chen, X. Yang and J. Sun, ”A Convolutional Neural Network-Based Chinese Text Detection Algorithm via Text Structure Modeling,” in IEEE Transactions on Multimedia, vol. 19, no. 3, pp. 506-518, March 2017, doi: 10.1109/TMM.2016.2625259.
17. W. He, X. -Y. Zhang, F. Yin and C. -L. Liu, ”Multi-Oriented and Multi-Lingual Scene Text Detection With Direct Regression, ” in IEEE Transactions on Image Processing, vol. 27, no. 11, pp. 5406-5419, Nov. 2018, doi: 10.1109/TIP.2018.2855399.
18. S. Shilaskar, S. Ghadge, M. Dhopade, J. Godle and M. Gulhane, ”English to Marathi Text Translation using Deep learning, ” 2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2022, pp. 1-5, doi: 10.1109/CONECCT55679.2022.9865781.
19. M. C. Anil, S. D. Shirbahadurkar and S. S. Shakil, ”A mapper and combiner based Marathi text to speech synthesis using English TTS Engine,” 2015 Annual IEEE India Conference (INDICON), New Delhi, India, 2015, pp. 1-5, doi: 10.1109/INDICON.2015.7443570.
20. T. Kano, S. Sakti and S. Nakamura, ”End-to-End Speech Translation With Transcoding by Multi-Task Learning for Distant Language Pairs,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 1342-1355, 2020, doi: 10.1109/TASLP.2020.2986886.
21. S. Nakamura et al., ”The ATR Multilingual Speech-to-Speech Translation System,” in IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, no. 2, pp. 365-376, March 2006, doi: 10.1109/TSA.2005.860774.
22. T. Kano, S. Sakti and S. Nakamura, ”End-to-End Speech Translation With Transcoding by Multi-Task Learning for Distant Language Pairs,” in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 1342-1355, 2020, doi: 10.1109/TASLP.2020.2986886.
23. R. T. sairaj and S. R. Balasundaram, ”Improving the Cognitive Levels of Automatic Generated Questions using Neuro-Fuzzy Approach in e-Assessment,” 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2020, pp. 454-458, doi: 10.1109/ICCCA49541.2020.9250716.
24. A. R. Singh, D. Bhardwaj, M. Dixit and L. Kumar, ”An Integrated Model for Text to Text, Image to Text and Audio to Text Linguistic Conversion using Machine Learning Approach,” 2023 6th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2023, pp. 1-7, doi: 10.1109/ISCON57294.2023.10112123.
25. A. Rasheed, N. Ali, B. Zafar, A. Shabbir, M. Sajid and M. T. Mahmood, ”Handwritten Urdu Characters and Digits Recognition Using Transfer Learning and Augmentation With AlexNet,” in IEEE Access, vol. 10, pp. 102629-102645, 2022, doi: 10.1109/ACCESS.2022.3208959.