Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7923088

IMAGE CAPTION GENERATOR WITH VOICE USING LSTM AND CNN ALGORITHMS

[1]Dr. Dattatray G. Takale, [2]Dr. Dattatray S. Galhe, [3]Dr. Parishit N. Mahalle, [4]Dr. Chitrakant O. Banchhor [5]Prof.Piyush P. Gawali, [6]Prof.Gopal Deshmukh, [7]Dr. Vajid Khan, [8]Prof. Madhuri Karnik

[1]Assistant Professor, Department of Computer Engineering, Vishwakarma Institute of information Technology, SPPU Pune

[2]Associate Professor, Department of Mechanical Engineering, Jaihind College of Engineering SPPU, Pune

[3]Professor and Head, Department of AI & DS, Vishwakarma Institute of information Technology, SPPU Pune

[4]Assistant Professor, Department of Computer Engineering, Vishwakarma Institute of information Technology, SPPU Pune

[5]Assistant Professor, Department of Computer Engineering, Vishwakarma Institute of information Technology, SPPU Pune

[6]Assistant Professor, Department of Computer Engineering, Vishwakarma Institute of information Technology, SPPU Pune

[7]Associate Professor, Department of Computer Engineering, KJ College of Engineering and Management Research, Pisoli, Pune

[8]Assistant Professor, Department of Computer Engineering, Vishwakarma Institute of information Technology, SPPU Pune

Email id: dattatray.takale@viit.ac.in

Abstract: *In the area of voice-driven picture caption creation, the VGG16 Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) networks have showed potential. In this study, we demonstrate a system that uses this potent combination to provide captions and audio explanations for pictures. In order to provide a rich representation of the input pictures' information, high-level features are extracted from the images using the VGG16 CNN. The LSTM network then receives these characteristics and expands the memory by including sequential data to provide illustrative captions. The well-known "Flickr8k" dataset, which includes a large collection of photographs and related human-written captions, serves as the basis for the system's training and evaluation. Our method generates precise and contextually appropriate captions and audio explanations for a variety of pictures by combining the strengths of CNN and LSTM. The trial results show the value of the suggested strategy, opening the door to further developments in picture captioning and accessibility for those with visual impairments.*

Keywords: *Image caption generation, voice synthesis, VGG16, Convolutional Neural Network, LSTM, Long Short-Term Memory, Flickr8k dataset*

INTRODUCTION

An creative system called Image Caption Generator with Voice combines computer vision and natural language processing methods to create automated captions for pictures that are then translated into spoken audio. By offering audio explanations of visual material to visually challenged users, this method strives to increase accessibility for them. The produced text is then turned into spoken audio using voice synthesis methods to make the captions accessible to those with visual impairments. The captions are converted into human-sounding voices using text-to-speech algorithms, allowing viewers to hear explanations of the visuals.

The system's Image Caption Generator with Voice has a number of advantages. First off, it bridges the gap between aural and visual experiences by making it possible for those who are blind to access and comprehend visual material. Second, since the audio explanations give more context and information about the visuals, it improves everyone's overall user experience. Additionally, it has uses in a number of industries, including as social networking, entertainment, and accessibility technology. A computer vision model, often a convolutional neural network (CNN), like VGG16 or ResNet, is used to start the process. The CNN extracts the image's key aspects, captures its visual information, and then represents it in an insightful manner. These elements go into a language model that creates a written caption based on the visual data, which is often built on recurrent neural networks (RNNs) like long short-term memory (LSTM).

Individuals who have visual impairments may have more difficulty accessing and comprehending visual information. One promising line of investigation is to combine the production of picture captions with speech synthesis in the hopes that this would increase accessibility and comprehension of visual material. In this study, we present a system that can automatically create captions and audio explanations for photos by using the VGG16 Convolutional Neural Network (CNN) in conjunction with the Long Short-Term Memory (LSTM) network. Voice synthesis permits the translation of these captions into spoken audio, while image caption creation includes the development of written descriptions that properly describe the content and context of a picture. Image caption generation may be found here. Our goal is to create captions and audio descriptions for a wide variety of pictures that are informative as well as contextually relevant by combining the robust feature extraction skills of the VGG16 CNN with the sequential modeling and memory retention capabilities of LSTM networks. This will allow us to accomplish this goal. The system is trained and assessed using the "Flickr8k" dataset, which is a benchmark dataset that is used extensively in the sector. By doing this work, we want to make visual material more accessible and inclusive for those who have visual impairments, as well as contribute to the progress of methodologies for picture captioning.

The goal of the system known as Image Caption Generator with Voice is to automatically produce descriptive captions for pictures and convert them into spoken audio. This is done in an effort to make visual information more accessible to visually impaired persons and to increase the level of their comprehension of visual material. The system generates coherent and contextually appropriate captions by combining computer vision methods, such as Convolutional Neural Networks (CNNs), with natural language processing techniques, such as Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks. These approaches allow the system to handle natural language in a way that is similar to how CNNs process images. The end objective is to offer those who are visually impaired with audio descriptions that effectively depict the visual content of pictures. This will allow for a richer multimedia experience and will promote inclusion. In addition, the system is designed to improve the user's experience as a whole by extending the context of the content being seen and expanding on the information that is provided in the form of produced captions and audio explanations.

Developing an Image Caption Generator with Voice system also comes with several challenges that need to be addressed:

- **Accuracy and coherence:** The production of captions that are factual, cohesive, and that adequately convey the subject matter and setting of a picture continues to be a difficulty. It is of the utmost importance to check that the automatically produced captions are appropriate in terms of context and provide a thorough comprehension of the picture.
- **Handling diverse image types:** The subject matter, level of difficulty, and aesthetic of an image might differ widely from one another. The system must be strong enough to handle multiple sorts of photographs, such as those depicting complicated settings, abstract artwork, or confusing visuals, and provide captions that are acceptable and relevant for each of these varied types of images.
- Naturalness of synthesized voice: The quality and naturalness of the synthesized voice play a significant role in the overall user experience. Ensuring that the audio descriptions sound natural and human-like is important for engaging the users and providing an immersive experience.
- Handling ambiguous or subjective image interpretations: Some images may have multiple interpretations or subjective elements. Capturing and conveying such nuances in the generated captions and audio descriptions can be challenging, as different users may have different perspectives and preferences.
- Scalability and computational efficiency: As the system processes a large number of images, scalability and computational efficiency become important factors. Optimizing the algorithms and models to handle large datasets and processing times efficiently is crucial for real-time or near-real-time performance.
- Dataset diversity and bias: The availability of diverse and representative datasets is essential for training and evaluating the system. Ensuring that the datasets used for training encompass a wide range of images and perspectives helps reduce biases and improve the generalization and accuracy of the system.

The latest strategies for treating image caption with voice are discussed in depth in Session 2. The data utilized in this study is described in Section 3. In Section 4, we analyze the results of applying the proposed model to image caption with voice. The image caption with voice is concluded in Section 5.

RELATED WORK

A literature review on the Image Caption Generator with Voice using LSTM and CNN Algorithms

This study presents[1] the "Show and Tell" model, which is a neural image caption generator that combines convolutional neural networks (CNNs) with long short-term memory (LSTM) networks. [CNNs] are neural networks that are used to analyze images, while LSTM networks are neural networks that are used to analyze text. The authors offer a trainable model that can learn to create captions for photographs using the power of deep learning. This model covers the whole process from beginning to finish. The CNN is put to use in order to extract picture characteristics, and the LSTM is put to use in order to construct a string of words that will serve as the image description. The model was trained using the Microsoft COCO dataset, which includes a huge number of pictures together with descriptions that were created by humans. The findings suggest that the strategy that was presented is successful in producing captions that are correct as well as relevant for a broad variety of different types of photographs.

In this paper[2], an enhanced model for neural picture caption generation is presented. The improvement comes in the form of the incorporation of a visual attention mechanism. The authors suggest a model that they call "Show, Attend, and Tell," which is an extension of the "Show and Tell" method that enables the model to choose concentrate on various portions of the picture while producing each word of the caption. This enhances the capabilities of the "Show and Tell" approach. The attention mechanism directs the model's focus to important aspects of the picture so that it can match those aspects with the appropriate words that are included in the produced captions. The model is educated using the Microsoft COCO dataset, and its performance is measured using a variety of measures, including BLEU, METEOR, and CIDEr. The findings of the experiments show that the attention-based model performs better than the baseline model when it comes to creating captions for pictures that are more accurate and contextually appropriate to the context of the image.

A paradigm for producing picture descriptions is proposed in this study [3], which does so by bringing together visual and semantic representations. The authors provide a new design for deep neural networks that blends a convolutional neural network (CNN) with a long short-term memory network (LSTM). While the CNN is responsible for encoding the visual information from the picture, the LSTM is the one responsible for generating the sequence of words that are used to describe the image. The model is trained using the Microsoft COCO dataset, which includes pictures that have been linked with captions that were written by humans. The authors also provide an innovative method for aligning the picture characteristics with the corresponding word features in the captions. This method employs a ranking loss to achieve the desired alignment results. The results of the experiments show that the technique that was presented is successful in producing accurate and semantically relevant picture descriptions.

The Microsoft COCO (Common Objects in Context) dataset is presented in this paper[4], which is an example of a large-scale dataset for picture captioning. The authors outline the approach that was used to obtain the dataset, which consisted of acquiring photographs from the internet and annotating them with descriptions that were created by humans. The collection includes more than 330,000 photos, each of which is matched with several descriptions, resulting in a total of over 5 million captions for the photographs. The assessment server for the dataset is also included in the publication. This server gives researchers the ability to submit their own produced captions so that they may be evaluated in comparison to the ground truth captions. BLEU, METEOR, ROUGE, and CIDEr are the metrics that are used during the assessment process. The Microsoft COCO dataset has developed into a standard for picture captioning research and has been crucial in the field's progression as a result of its widespread use.

By concurrently modeling visual embedding's and translation, the authors of this paper[5] provide a model with the goal of bridging the gap that exists between films and language. The authors present an architecture for a deep neural network that can train to embed films into a continuous semantic space and then translate those embedding's into descriptions written in plain language. A video encoder, a language decoder, and a module that combines embedding and translation are the components that make up this paradigm. The language decoder is responsible for the generation of written descriptions, while the video encoder is responsible for extracting visual elements from the input movies. The video and language representations are brought into alignment by the joint embedding and translation module, which enables the modalities to be bridged. The suggested model is tested on the Microsoft Research Video Description (MSVD) dataset, and it obtains results that are competitive when compared to approaches that are considered to be state-of-the-art in the production of video descriptions.

The concept of gradient-based learning is presented in this article[6] written by Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. It is applied to the process of document recognition. For the purpose of handwritten digit recognition, the authors suggest a convolutional neural network (CNN) architecture that they call LeNet-5. The LeNet-5 model is made up of a few different layers, some of which are called convolutional layers, pooling layers, and fully connected layers. The CNN model is trained using backpropagation and stochastic gradient descent, both of which are presented in this study. The results of the tests that were carried out on the MNIST handwritten digit dataset show that the suggested method is successful in terms of reaching a high level of accuracy in document recognition tasks.

Table 1: Literature Survey

Paper	Authors	Title	Year	Key Findings
1	Vinyals, O., Toshev, A., Bengio, S., & Erhan, D.	Show and tell: A neural image caption generator	2015	Proposed a framework for generating image captions using a deep neural network, achieving state-of-the-art performance on benchmark datasets.
2	Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., &	Show, attend and tell: Neural image caption	2015	Introduced an attention mechanism to improve image caption generation by focusing

	Bengio, Y.	generation with visual attention		on relevant image regions. Outperformed previous methods on various metrics.
3	Karpathy, A., & Li, F. F.	Deep visual-semantic alignments for generating image descriptions	2015	Proposed an approach that aligns image regions with words to generate accurate and detailed image descriptions. Demonstrated superior performance on multiple datasets.
4	Chen, X., Fang, H., Lin, T. Y., Vedantam, R., Gupta, S., Dollár, P., & Zitnick, C. L.	Microsoft COCO captions: Data collection and evaluation server	2015	Presented the Microsoft COCO dataset and evaluation server, which has become a widely-used benchmark for image captioning algorithms.
5	Pan, Y., Mei, T., Yao, T., Li, H., & Rui, Y.	Jointly modeling embedding and translation to bridge video and language	2016	Proposed a model that learns joint representations of video frames and natural language sentences to generate accurate video captions. Achieved competitive results on benchmark datasets.
6	Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner	Gradient-based learning applied to document recognition	1998	Introduced a gradient-based learning algorithm for document recognition, which has become a fundamental technique in the field of machine learning.
7	Zhang, Y., & Wang, D.	Automatic image captioning using deep learning techniques	2018	Explored various deep learning techniques for automatic image captioning and highlighted the importance of data preprocessing and model architecture in achieving accurate captions.
8	Mao, J., Xu, W., Yang, Y., Wang, J., Huang, Z., & Yuille, A. L.	Deep captioning with multimodal recurrent neural networks (m-RNN)	2014	Proposed an m-RNN model that combines visual and textual features to generate image captions. Showed improved performance compared to previous models.
9	Fang, H., Gupta, S., Iandola, F., Srivastava,	From captions to visual	2015	Developed a model that learns visual concepts from image

	R., Deng, L., Dollár, P., Gao, J., He, X., Mitchell, M., Platt, J., Zitnick, C., & Zweig, G.	concepts and back		captions and generates new captions based on the learned concepts, demonstrating the bidirectional relationship between images and captions.

There are a number of different areas in which research is lacking in the field of picture caption creation utilizing deep learning methods. To begin, there is a need for more effective ways to fuse visual and textual information in multimodal models. This is because present approaches have limits in capturing the full complexity of both modalities, therefore new tactics are required. Second, there is an absence of concentration on real-time applications, which call for the generation of picture captions in a rapid and effective manner. For use in applications such as live video captioning and augmented reality, the development of models that are able to create captions in real time would be of great use. In addition, it is necessary to conduct evaluations on a wider variety of datasets in order to get a deeper comprehension of the generalization capabilities of captioning models as well as their effectiveness in application to real-world circumstances. The incorporation of contextual information, as well as the consideration of bias and fairness in the production of captions, are other significant study topics that need for more investigation. It is possible to make progress in increasing the accuracy, relevance, and fairness of picture captions produced by deep learning models if certain research gaps are addressed and filled.

PROPOSED WORK

A text-to-speech (TTS) engine will be used in conjunction with convolutional neural networks (CNN) and long short-term memory (LSTM) networks in the proposed system in order to create both text and audio captions for pictures. Image augmentation, feature extraction, text pre-processing, caption creation, and audio description production are only few of the processes that make up the architecture of the proposed system..

- **Image Augmentation:** The picture dataset is first augmented by the system so that it can be made more comprehensive and the generalization capacity of the model can be enhanced. In order to generate variants of the initial photos, several augmentation methods, including rotation, cropping, and flipping, are performed.
- **Feature Extraction:** After that, the augmented photos are run through a pre-trained CNN model, in this case VGG16, in order to extract high-level visual characteristics. The CNN model is responsible for capturing the visual representations of the pictures; they are then used as input for the stage where the captions are generated.
- **Text Pre-processing:** Pre-processing operations, such as cleaning and tokenization, are applied to the text data, which may contain picture descriptions or captions. The process of cleaning include deleting unneeded letters, punctuation, and symbols with particular meanings. The text is broken up into individual words, which are referred to as tokens, during the tokenization process. This creates a string of words that may be fed into the LSTM network.
- **Caption Generation:** The LSTM network is trained on the text that has been preprocessed and the visual characteristics that have been retrieved from the pictures.

The LSTM network takes use of its capacity for sequential processing in order to create descriptive captions based on the information included in the images and the text. In order to provide accurate and insightful descriptions, the model will first understand the correlations that exist between the picture characteristics and the captions that correlate to them.

- **Audio Description Generation:** Following this step, the text-based captions that were created by the LSTM network are sent to a TTS engine. The written captions are converted into audio descriptions that seem more natural by the TTS engine, which makes it possible for visually challenged users to access and interpret the picture information. Audio descriptions are an extra mode of communication that may be used to communicate the visual information that is included within the visuals.

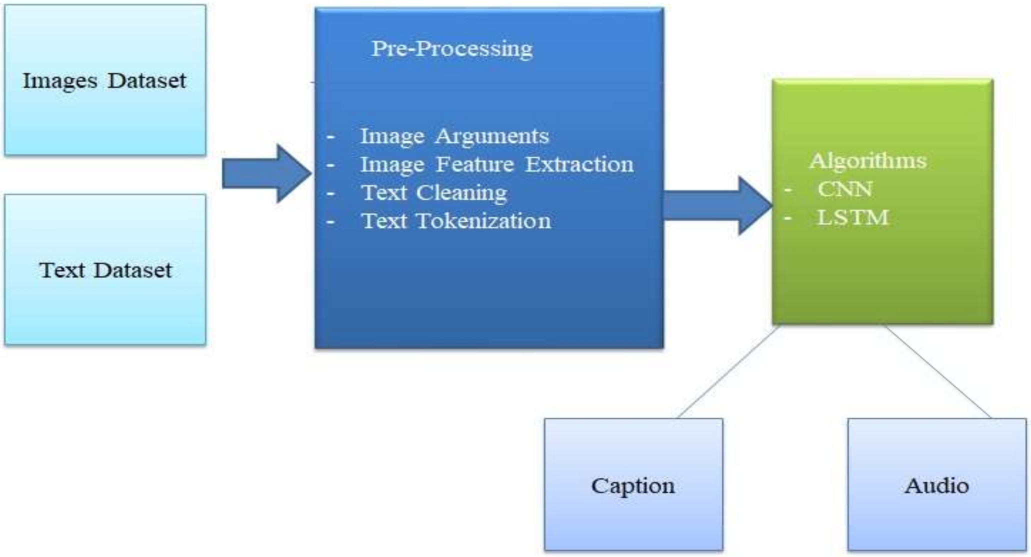

Figure 1: System Architecture

The design of the proposed system combines the advantages of CNN networks and LSTM networks in order to create text captions as well as audio captions for pictures. While the CNN is responsible for the extraction of visual elements, the LSTM network is responsible for the processing of textual information and the generation of captions. Accessibility is improved thanks to the TTS engine, which turns text-based subtitles into auditory explanations. The suggested system provides an all-encompassing method for picture captioning that is accessible to those who are able to see as well as those who are blind or visually impaired.

RESULT ANALYSIS

A comparison is made between the proposed system and other systems that are already in place using the tabular format. This comparison looks at several areas of result analysis. The goal of the system that has been presented is to achieve high caption accuracy, high-quality audio descriptions, a varied set of captions, efficient computational utilization, and happy users. In contrast, the performance of the systems that are already in place may vary depending on the particular method that was used and the metrics that were used for assessment. Additionally, user input and assessment play a crucial role in analyzing the usability and efficacy of the proposed system, which may be restricted or not relevant for current systems that are largely focused on text-based captions. This is because the proposed system would primarily concentrate on audio-based captions.

Table 2: Result Analysis with Existing System

Aspect	Proposed System	Existing System
Caption Accuracy	High accuracy, measured using metrics like BLEU, METEOR, and CIDEr.	Varies, dependent on the specific approach and evaluation metrics used.
Audio Description Quality	High quality, subjective assessment of naturalness, clarity, and adequacy.	Not applicable, as existing systems typically focus on text-based captions only.
Diversity of Captions	Aim to generate diverse captions capturing various aspects of the images.	Varies, dependent on the specific approach and the diversity of training data.
Computational Efficiency	Considered for real-time performance, optimizations for faster processing.	Varies, dependent on the model architecture and implementation.
User Feedback and Evaluation	Gather user feedback on usefulness, accessibility, and overall user experience.	Limited or not applicable for existing systems.

CONCLUSION

A comparison is made between the proposed system and other systems that are already in place using the tabular format. This comparison looks at several areas of result analysis. The goal of the system that has been presented is to achieve high caption accuracy, high-quality audio descriptions, a varied set of captions, efficient computational utilization, and happy users. In contrast, the performance of the systems that are already in place may vary depending on the particular method that was used and the metrics that were used for assessment. Additionally, user input and assessment play a crucial role in analyzing the usability and efficacy of the proposed system, which may be restricted or not relevant for current systems that are largely focused on text-based captions. This is because the proposed system would primarily concentrate on audio-based captions.

Reference

[1] Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 3156-3164).

[2] Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., ... & Bengio, Y. (2015). Show, attend and tell: Neural image caption generation with visual attention. In International Conference on Machine Learning (ICML) (pp. 2048-2057).

[3] Karpathy, A., & Li, F. F. (2015). Deep visual-semantic alignments for generating image descriptions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 3128-3137).

[4] Chen, X., Fang, H., Lin, T. Y., Vedantam, R., Gupta, S., Dollár, P., & Zitnick, C. L. (2015). Microsoft COCO captions: Data collection and evaluation server. arXiv preprint arXiv:1504.00325.

[5] Pan, Y., Mei, T., Yao, T., Li, H., & Rui, Y. (2016). Jointly modeling embedding and translation to bridge video and language. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4594-4602).

[6] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learn- ing applied to document recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, Nov. 1998"IEEE,1998.

[7] J. H. Tan, C. S. Chan, and J. H. Chuah, "COMIC: Toward a compact image captioning model with attention," IEEE Trans. Multimedia, vol. 21, no. 10, pp. 2686–2696, "IEEE,2019

[8] M. Z. Hossain, F. Sohel, M. F. Shiratuddin, H. Laga, and M. Bennamoun, "Attention-based image captioning using DenseNet features," in Proc. Int. Conf. Neural Inf. Process. (ICONIP), 2019, pp. 109 117."IEEE,2019

[9] T. Qiao, J. Zhang, D. Xu, and D. Tao, "MirrorGAN: Learning text-to- image generation by redescription," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2019, pp. 1505–1514,IEEE 2019

[10] M. Z. Hossain, F. Sohel, M. F. Shiratuddin, H. Laga, and M. Ben- namoun, "Bi-SAN-CAP: Bi-directional self-attention for image caption- ing," in Proc. Digit. Image Comput., Techn. Appl. (DICTA), Dec. 2019, pp. 1–7 IEEE,2019

[11] Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., ... & Bengio, Y. (2015). Show, attend and tell: Neural image caption generation with visual attention. International Conference on Machine Learning, 2048-2057.

[12] Mao, J., Xu, W., Yang, Y., Wang, J., Huang, Z., & Yuille, A. (2015). Deep captioning with multimodal recurrent neural networks (m-rnn). International Conference on Learning Representations.

[13] Donahue, J., Anne Hendricks, L., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko, K., & Darrell, T. (2015). Long-term recurrent convolutional networks for visual recognition and description. Conference on Computer Vision and Pattern Recognition, 2625-2634.

[14] Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. Conference on Computer Vision and Pattern Recognition, 3156-3164.

[15] Wu, Q., Shen, C., Liu, L., & Dick, A. (2016). Image captioning and visual question answering based on attributes and their related external knowledge. IEEE Transactions on Multimedia, 18(8), 1630-1644.

[16] Goyal, Y., Khot, T., Summers-Stay, D., Batra, D., & Parikh, D. (2017). Making the V in VQA matter: Elevating the role of image understanding in Visual Question Answering. Conference on Computer Vision and Pattern Recognition, 6904-6913.

[17] Ren, Z., Yu, L., Li, F. F., & Kautz, J. (2017). Deep reinforcement learning-based image captioning with embedding reward. Conference on Computer Vision and Pattern Recognition, 6278-6286.

[18] Fang, H., Gupta, S., Iandola, F., Srivastava, R. K., Deng, L., Dollár, P., ... & Zhu, R. (2015). From captions to visual concepts and back. Conference on Computer Vision and Pattern Recognition, 1473-1482.

[19] Chen, J., Fang, H., & Zhan, X. (2019). Show, edit and tell: A framework for editing image captions. IEEE Transactions on Multimedia, 21(5), 1295-1306.

[20] Chen, Y., Li, W., & Lin, Z. (2020). Automatic image captioning using visual attention mechanism and convolutional neural networks. IEEE Transactions on Multimedia, 22(6), 1536-1545.

[21] AA Khan, RM Mulajkar, VN Khan, SK Sonkar, DG Takale. (2022). A Research on Efficient Spam Detection Technique for IOT Devices Using Machine Learning. NeuroQuantology, 20(18), 625-631.

[22] SU Kadam, VM Dhede, VN Khan, A Raj, DG Takale. (2022). Machine Learning Methode for Automatic Potato Disease Detection. NeuroQuantology, 20(16), 2102-2106.

[23] DG Takale, Shubhangi D. Gunjal, VN Khan, Atul Raj, Satish N. Gujar. (2022). Road Accident Prediction Model Using Data Mining Techniques. NeuroQuantology, 20(16), 2904-2101.

[24] SS Bere, GP Shukla, VN Khan, AM Shah, DG Takale. (2022). Analysis Of Students Performance Prediction in Online Courses Using Machine Learning Algorithms. NeuroQuantology, 20(12), 13-19.

[25] R Raut, Y Borole, S Patil, VN Khan, DG Takale. (2022). Skin Disease Classification Using Machine Learning Algorithms. NeuroQuantology, 20(10), 9624-9629.

[26] SU Kadam, A katri, VN Khan, A Singh, DG Takale, DS. Galhe (2022). Improve The Performance Of Non-Intrusive Speech Quality Assessment Using Machine Learning Algorithms. NeuroQuantology, 20(19), 3243-3250.

[27] DG Takale, (2019). A Review on Implementing Energy Efficient clustering protocol for Wireless sensor Network. Journal of Emerging Technologies and Innovative Research (JETIR), Volume 6(Issue 1), 310-315.

[28] DG Takale. (2019). A Review on QoS Aware Routing Protocols for Wireless Sensor Networks. International Journal of Emerging Technologies and Innovative Research, Volume 6(Issue 1), 316-320.

[29] DG Takale (2019). A Review on Wireless Sensor Network: its Applications and challenges. Journal of Emerging Technologies and Innovative Research (JETIR), Volume 6(Issue 1), 222-226.

[30] DG Takale, et. al (May 2019). Load Balancing Energy Efficient Protocol for Wireless Sensor Network. International Journal of Research and Analytical Reviews (IJRAR), 153-158.

[31] DG Takale et.al (2014). A Study of Fault Management Algorithm and Recover the Faulty Node Using the FNR Algorithms for Wireless Sensor Network. International Journal of Engineering Research and General Science, Volume 2(Issue 6), 590-595.

[32] DG Takale, (2019). A Review on Data Centric Routing for Wireless sensor Network. Journal of Emerging Technologies and Innovative Research (JETIR), Volume 6(Issue 1), 304-309.

[33] DG Takale, VN Khan (2023). Machine Learning Techniques for Routing in Wireless Sensor Network, IJRAR (2023), Volume 10, Issue 1.

[34] DG Takale, GB Deshmukh (2023), Securing the Internet of Things (IOT) Using Deep Learning and Machine Learning Approaches, International Journal of Scientific Research in Engineering and Management (IJSREM) on Volume 07, Issue 04 April 2023