IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 3, September 2024, pp. 2971~2977
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i3.pp2971-2977

Encoder-decoder approach for describing health of cauliflower plant in multiple languages

Parag Jayant Mondhe[1], Manisha P. Satone[1], Namrata N. Wasatkar[2]
[1]Department of Electronics and Telecommunication Engineering, Matoshri College of Engineering and Research Centre, Nashik, India
[2]Department of Computer Engineering, Vishwakarma Institute of Information Technology, Pune, India

Article Info

Article history:

Received Jun 21, 2023
Revised Jan 24, 2024
Accepted Feb 10, 2024

Keywords:

Aerial images
Cauliflower plant
Description generation
Encoder decoder approach
Multi language captions

ABSTRACT

Physically examining each plant to determine its state of health and determining the disease if plant is affected due to it, is challenging. The encoder - decoder approach is proposed for describing health of cauliflower plant in English, Hindi, and Marathi languages from aerial images. Experiments are performed with different convolutional neural network (CNN) models and long short-term memory (LSTM) combinations. The multilanguage cauliflower captions dataset (MCCD) is developed to evaluate the performance of the model. The dataset contains 1213 images where each image is described in 3 different languages. The dataset contains images of cauliflower plant affected due to bacterial spot rot, black rot, and downy mildew diseases. It also contains images of healthy plant. The objective metrics such as bilingual evaluation understudy (BLEU) scores and subjective criteria are used to decide the quality of the generated description.

Corresponding Author:

Parag Jayant Mondhe
Department of Electronics and Telecommunication Engineering
Matoshri College of Engineering and Research Centre
Eklahareshivar, Near Odhagaon, Opp Nashik-Aurangabad Highway, Nashik, Maharashtra 422105, India
Email: mondheparag@gmail.com

1. INTRODUCTION

The manual methods of crop monitoring are labour intensive, time consuming and may results in grossly inaccurate estimates. By automating the process of monitoring of crop health these challenges can be addressed. This paper proposes automation of crop monitoring process. The encoder-decoder approach is proposed for describing health of cauliflower plant in English, Hindi, and Marathi languages. The aerial images of farmland can be taken using unmanned aerial vehicle (UAV). Later, these images will be analysed by machine learning algorithm to describe health of the crop using captions in multiple languages.

The Hindi language is official language of Government of India and among top 5 globally spoken languages [1]. The third most widely spoken language in India is Marathi, which is also the official language of the state of Maharashtra [2]. Cauliflower is a member of the cruciferous family and has nutritional value. It contains vitamins, nutrients, fiber, and antioxidants [3], [4]. The Figure 1 shows the global production/yield quantities of cauliflowers and broccoli from year 1994 to 2020 [5].

The Figure 2 shows top cauliflower and broccoli producing countries during 1994 to 2021 [5]. During this period India was its second largest producer. The cauliflower plant is commonly affected by bacterial spot rot, black rot, and downy mildew diseases. Previously, researchers concentrated primarily on recognizing cauliflower disease from images [6]–[10] and did not explore generating descriptions about its health from aerial images in many languages. Three types of methodologies are utilized in the process of generating descriptions: encoder-decoder approaches [11]–[16], picture retrieval [17], and object recognition

[18]. In comparison to other approaches, the encoder-decoder approach produced superior results [19]. As a result, an encoder-decoder approach is proposed here for generating descriptions in multiple languages.

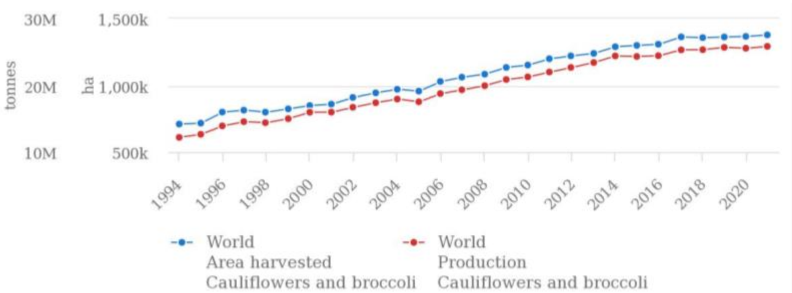

Figure 1. Global production/yield quantities of cauliflowers and broccoli [5]

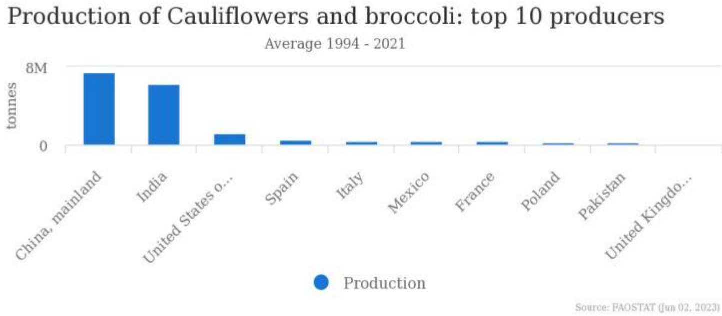

Figure 2. Top Cauliflower and broccoli producing countries [5]

2. PROPOSED ENCODER–DECODER APPROACH

The suggested encoder-decoder method for using an aerial image to express the health of a cauliflower plant in several languages is depicted in Figure 3. The images of farmland will be captured using UAV such as a drone. These images will act as in input to the encoder block.

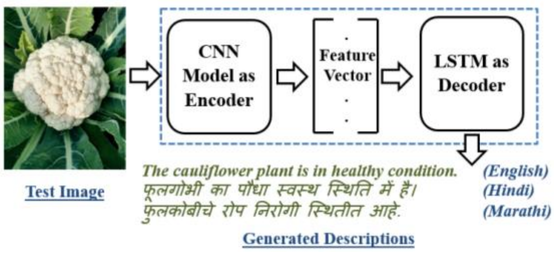

Figure 3. Proposed methodology for explaining health of cauliflower plant in multiple languages

Using aerial image of cauliflower fields as input, convolutional neural network (CNN) acts as an encoder in the model, extracting features from them. To prevent losing spatial information, the output of the CNN model's pooling layer is used instead of the last fully linked layer. The decoder is given the CNN-generated features as well as their descriptions. However, in the suggested method, the long short-term memory (LSTM) network serves as a decoder since it overcomes the recurrent neural networks (RNN) declining gradient problem [20].

During the training phase, the decoder learns how to provide a description of an image using its characteristics and previously produced words. In this method, each word is given a probability depending on its distinguishing characteristics and the preceding word. Figure 4 depicts the flowchart of proposed encoder-decoder approach. The same approach is followed to generate captions in other languages because we transform words to numbers in step 8 of the flowchart.

Experiments on feature extraction were carried out using several pre-trained CNN architectures, including visual geometry group 16 (VGG-16) [21], InceptionResNetV2 [22], and EfficientNetV2L [23]. The top-1 accuracy of VGG-16, InceptionResNetV2 and EfficientNetV2L CNN model is 71.30%, 80.30% and 85.70% respectively on ImageNet validation dataset [24]. The top-1 accuracy of the model indicates that the predicted label matches the target label. The top-5 accuracy of VGG-16, InceptionResNetV2, and EfficientNetV2L CNN model is 90.10%, 95.30%, and 97.50% respectively on ImageNet validation dataset [24]. The top-5 accuracy of the model indicates that the target label is one among the top 5 predicted label.

The VGG-16 is a CNN architecture that stands for VGG-16. VGG-16 is known for its depth because to its 16 layers, which include 13 convolutional layers and 3 fully linked layers. The design is made up of smaller 3×3 convolutional filters layered on top of each other, allowing for more detailed aspects of the input image to be captured. It employs a straightforward and consistent design paradigm in which convolutional layers are followed by max-pooling layers to lower spatial dimensions. The ImageNet dataset was used to train the VGG-16 model, which comprises millions of labelled images from diverse object categories.

A combination of the Inception and ResNet models is the InceptionResNetV2 model. The Inception architecture, upon which the InceptionResNetV2 model is based, combines convolutional layers with varying kernel sizes to collect information at various scales. EfficientNetV2L is an EfficientNet model family CNN model. The "V2L" in EfficientNetV2L stands for "vision to language". This model was created especially for multimodal activities that need the processing of textual and visual data.

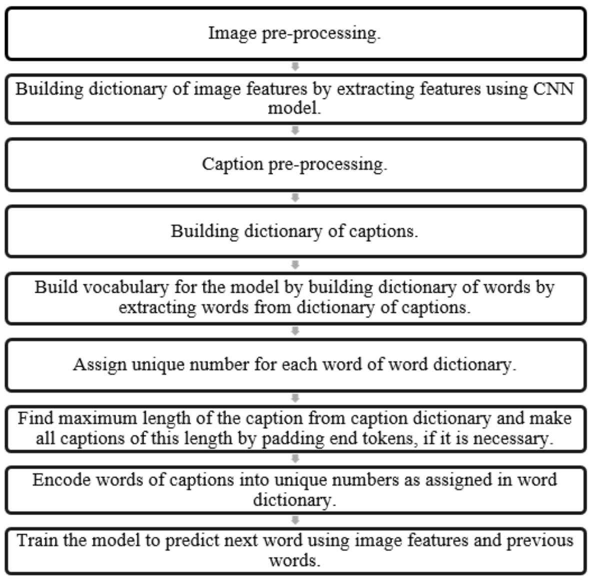

Figure 4. Flowchart for proposed approach

3. DATASET

The multilanguage cauliflower captions dataset (MCCD) is developed to evaluate the proposed model's performance. The dataset comprises 1213 colour images, each of which is accompanied by a caption written in English, Hindi, and Marathi language. The 656 images were captured in Bangladesh and provided in [25]. The remaining 557 images were obtained by the authors in 2023 on farmland in the Nashik district of Maharashtra State, India using a drone. The composition of the MCCD is shown in Table 1.

Table 1. Composition of MCCD

Category	Number of images		
	Images from Bangladesh [25]	Images captured in India	Total images
Bacterial spot rot	173	126	299
Black rot	100	158	258
Downy mildew	177	106	283
No disease/healthy plant	206	167	373
Total	656	557	1213

An agronomist captions each image with information in Marathi, Hindi, and English on the health of the cauliflower plant. Thus, there are 1213 captions in the dataset in Marathi, Hindi, and English. The dataset includes images of cauliflower plants infected by various diseases such as bacterial spot rot, black rot, and downy mildew, as well as images of healthy plants. The images were acquired from many sources and in various weather conditions, resulting in a diversified collection.

4. RESULTS AND DISCUSSIONS

4.1. Evaluation on objective metrics

The resulting caption quality of the proposed approach is validated using quantitative metrics bilingual evaluation understudy (BLEU). The BLEU metric is established in [26], and it uses a weighted average to compare various length phrase matches to the reference sentence. It counts the number of times an n-gram appears in the produced caption and the dataset's reference caption, where an n-gram is a collection of one or more ordered words. The BLEU is a precision-based score that ranges from 0 to 1, with a higher value suggesting a better match. For calculation of the score, the brevity penalty (BP) is calculated as (1):

$$BP = \begin{cases} 1, & if\ c > r \\ e^{\left(1-\frac{r}{c}\right)}, & if\ c \ll r \end{cases} \tag{1}$$

Where, c is the length of the generated caption and r is the reference caption length. The BLEU score is computed as (2):

$$BLEU = BP \cdot exp\left(\sum_{n=1}^{N} w_n \ log\ p_n\right) \tag{2}$$

Where, pn is modified n-gram precision, N is n-gram length, w_n is the positive weights and the sum of w_n is one.

For various N values, the BLEU number is computed. BLEU-1 makes use of a unigram precision value, but BLEU-2 makes use of a geometric sum of unigram and bigram precision. The BLEU-3 uses the geometric average of unigram, bigram, and trigram accuracy, whereas the BLEU-4 uses the geometric average of unigram, bigram, trigram, and four-gram precision. The Table 2 presents the BLEU metrics for several CNN models. From VGG-16 through InceptionResNetV2 to EfficientNetV2L, the quality of caption improves on all four BLEU measures. The BLEU score achieved for captions generated in multiple languages for specific model is also comparable.

Table 2. Results on objective metrics

CNN model	Language of caption	BLEU-1	BLEU-2	BLEU-3	BLEU-4
VGG-16	English	0.79	0.76	0.73	0.69
	Hindi	0.78	0.74	0.72	0.69
	Marathi	0.78	0.75	0.73	0.68
InceptionResNetV2	English	0.86	0.81	0.77	0.74
	Hindi	0.86	0.82	0.77	0.73
	Marathi	0.85	0.81	0.76	0.72
EfficientNetV2L	English	0.90	0.84	0.80	0.77
	Hindi	0.90	0.83	0.79	0.75
	Marathi	0.89	0.82	0.80	0.75

4.2. Evaluation on subjective criteria

To determine the quality of the caption, the BLEU score just contrasts the generated caption with the reference captions. An objective agronomist thoroughly verifies the generated description in subjective judgement. Three categories—correct, partially correct, and incorrect—are created from the generated caption and are based on quality. The findings on subjective criteria for several CNN models are displayed in Table 3. Overall, caption quality is better from VGG-16 to InceptionResNetV2 to EfficientNetV2L.

Figure 5 compares the performance of several CNN models based on subjective criteria. Similar to objective assessments, the performance of various models for numerous languages is similar. The InceptionResNetV2 and EfficientNetV2L models beat the VGG-16 model, which is deemed shallow because to its low number of layers.

The Figure 6 shows results for EfficientNetV2L CNN model while Figure 7 shows the results for InceptionResNetV2 CNN model. The Figure 6 contains test image of cauliflower plant which is affected due to bacterial spot rot disease. The Figure 7 contains test image of cauliflower plant which is in healthy condition.

If the generated description accurately portrays the health of the cauliflower plant with no grammatical errors, it falls into the correct category. The caption generated in both English and Hindi language for test image provided in Figure 6 is categorized as correct. Similarly, the caption generated in English language for test image shown in Figure 7 is also categorized in correct category. If the description is erroneous, linguistically faulty, or useless, it is placed in the incorrect category. The caption generated in Marathi language for test image provided in Figure 7 falls into this category as the generated caption is meaningless.

Table 3. Results on subjective criteria

CNN model	Language of caption	Correct caption (%)	Partially correct caption (%)	Incorrect caption (%)
VGG-16	English	77.00	15.11	7.89
	Hindi	76.08	15.13	8.79
	Marathi	76.03	14.09	9.88
InceptionResNetV2	English	84.19	11.03	4.78
	Hindi	85.16	8.87	5.97
	Marathi	83.97	10.11	5.92
EfficientNetV2L	English	87.60	10.81	1.59
	Hindi	86.97	11.94	1.09
	Marathi	86.97	11.02	2.01

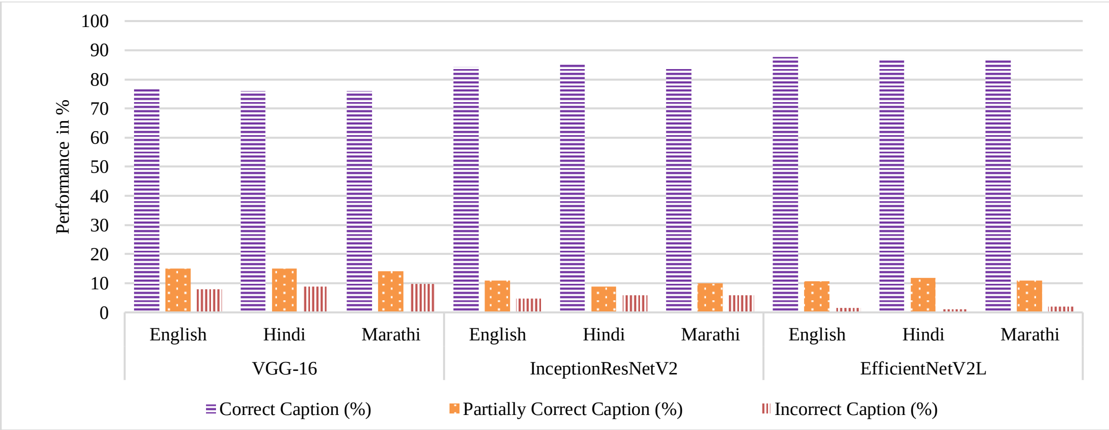

Figure 5. Comparison of results on subjective criteria

Language	Reference caption	Generated caption	Category of generated caption
English	Bacterial spot rot has infected the head of a cauliflower plant.	Bacterial spot rot has infected the head of a cauliflower plant.	Correct
Hindi	फूलगोभी के फूल जीवाणु रोग से संक्रमित हो गए हैं।	फूलगोभी के फूल जीवाणु रोग से संक्रमित हो गए हैं।	Correct
Marathi	फुलकोबीच्या फुलोऱ्याला जिवाणूजन्य आजाराचा संसर्ग झाला आहे.	फुलकोबीच्या फुलोऱ्याला संसर्ग झाला आहे.	Partially correct

Figure 6. Results for EfficientNetV2L CNN model

Language	Reference caption	Generated caption	Category of generated caption
English	A healthy and fresh cauliflower crop is growing at a reasonable rate.	A healthy and fresh cauliflower crop is growing at a reasonable rate.	Correct
Hindi	ताजा और स्वस्थ फूलगोभी की फसल अच्छी गति से बढ़ रही है।	स्वस्थ फूलगोभी की फसल गति से बढ़ रही है।	Partially correct
Marathi	ताजे आणि निरोगी फुलकोबीचे पीक योग्य गतीने वाढत आहे.	ताजे पीक आहे.	Incorrect

Figure 7. Results for InceptionResNetV2 CNN model

A statement that accurately characterizes the plant's health but does not give specifics is classified as partially correct. The caption generated in Marathi language for test image provided in Figure 6 is categorized as partially correct as caption indicated that cauliflower plant is affected but fails to provide details of the disease. Similarly, the caption generated in Hindi language for test image shown in Figure 7 is also categorized in same category as it hides other details of healthy plant.

5. CONCLUSIONS

The encoder-decoder approach is proposed to generate captions in multiple languages to describe health of cauliflower plant from aerial images. The experiments are performed with CNN models such as VGG-16, InceptionResNetV2, EfficientNetV2L, and LSTM combinations. On both BLEU score and subjective criteria EfficientNetV2L-LSTM combination has provided superior results. The InceptionResNetV2 and EfficientNetV2L have performed substantially well than VGG-16 as they have more layers which results in better feature extraction. The captions generated in various languages were of comparable quality.

REFERENCES

[1] Ministry of Electronics and Information Technology India, "Internet of things," *MeitY*, 2016. [Online]. Available: https://www.meity.gov.in/content/internet-things

[2] "Maharashtra state government." [Online]. Available: https://www.maharashtra.gov.in/

[3] J. Wang *et al.*, "A comparative study on the nutrients, mineral elements, and antioxidant compounds in different types of cruciferous vegetables," *Agronomy*, vol. 12, no. 12, 2022, doi: 10.3390/agronomy12123121.

[4] "United States of America (USA), Department of Agricultural Research Service," *USDA*. [Online]. Available: https://fdc.nal.usda.gov/

[5] FAO, "Food and agriculture organization of the United Nations," *International Organization*, 1947. [Online]. Available: https://www.fao.org

[6] T. Y. Orin, M. U. Mojumdar, S. M. T. Siddiquee, and N. R. Chakraborty, "Cauliflower leaf disease detection using computerized techniques," *2021 IEEE 6th International Conference on Computing, Communication and Automation, ICCCA 2021*. IEEE, pp. 730–733, 2021, doi: 10.1109/ICCCA52192.2021.9666437.

[7] S. K. Maria, S. S. Taki, M. J. Mia, A. A. Biswas, A. Majumder, and F. Hasan, "Cauliflower disease recognition using machine learning and transfer learning," *Smart Innovation, Systems and Technologies*, vol. 235. Springer Singapore, pp. 359–375, 2022, doi: 10.1007/978-981-16-2877-1_33.

[8] M. A. Malek, S. S. Reya, N. Zahan, M. Z. Hasan, and M. S. Uddin, "Deep learning-based cauliflower disease classification," *Algorithms for Intelligent Systems*. Springer Singapore, pp. 171–186, 2022, doi: 10.1007/978-981-16-9991-7_11.

[9] S. Kashyap, T. Thaware, S. R. Sahu, and K. M. Rao, "Multi-crop leaf disease detection using deep learning methods," *INDICON 2022 - 2022 IEEE 19th India Council International Conference*. IEEE, 2022, doi: 10.1109/INDICON56171.2022.10040099.

[10] K. P. A. Rani and S. Gowrishankar, "Pathogen-based classification of plant diseases: a deep transfer learning approach for intelligent support systems," *IEEE Access*, vol. 11, pp. 64476–64493, 2023, doi: 10.1109/ACCESS.2023.3284680.

[11] B. Qu, X. Li, D. Tao, and X. Lu, "Deep semantic understanding of high resolution remote sensing image," *2016 International Conference on Computer, Information and Telecommunication Systems*. IEEE, 2016, doi: 10.1109/CITS.2016.7546397.

[12] X. Zhang, X. Li, J. An, L. Gao, B. Hou, and C. Li, "Natural language description of remote sensing images based on deep learning," *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2017, pp. 4798–4801, 2017, doi: 10.1109/IGARSS.2017.8128075.

[13] X. Zhang, Q. Wang, S. Chen, and X. Li, "Multi-scale cropping mechanism for remote sensing image captioning," *International*

Geoscience and Remote Sensing Symposium (IGARSS), vol. 2019, pp. 10039–10042, 2019, doi: 10.1109/IGARSS.2019.8900503.
[14] X. Lu, B. Wang, X. Zheng, and X. Li, "Exploring models and data for remote sensing image caption generation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, pp. 2183–2195, 2018, doi: 10.1109/TGRS.2017.2776321.
[15] S. C. Kumar, M. Hemalatha, S. B. Narayan, and P. Nandhini, "Region driven remote sensing image captioning," *Procedia Computer Science*, vol. 165, pp. 32–40, 2019, doi: 10.1016/j.procs.2020.01.067.
[16] G. Hoxha and F. Melgani, "A novel SVM-based decoder for remote sensing image captioning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022, doi: 10.1109/TGRS.2021.3105004.
[17] B. Wang, X. Lu, X. Zheng, and X. Li, "Semantic descriptions of high-resolution remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 8, pp. 1274–1278, 2019, doi: 10.1109/LGRS.2019.2893772.
[18] Z. Shi and Z. Zou, "Can a machine generate humanlike language descriptions for a remote sensing image?," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3623–3634, 2017, doi: 10.1109/TGRS.2017.2677464.
[19] P. J. Mondhe, M. P. Satone, and G. K. Kharate, "Automatic caption generation for aerial images: a survey," *Indonesian Journal of Electrical Engineering and Informatics*, vol. 11, no. 1, pp. 267–285, 2023, doi: 10.52549/ijeei.v11i1.4342.
[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd International Conference on Learning Representations, ICLR 2015-Conference Track Proceedings*, pp. 1-14, 2015.
[22] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, vol. 31, no. 1, pp. 4278–4284, 2017, doi: 10.1609/aaai.v31i1.11231.
[23] M. Tan and Q. V. Le, "EfficientNetV2: smaller models and faster training," in *Proceedings of Machine Learning Research*, 2021, vol. 139, pp. 10096–10106.
[24] "Simple. Flexible. Powerful" *Keras.* [Online]. Available: https://keras.io
[25] A. Rajbongshi, U. S. SAra, R. Shakil, B. Akter, and M. S. Uddin, "VegNet: an extensive dataset of cauliflower images to recognize the diseases using machine learning and deep learning models," *Mendeley Data*, version 3, 2022, doi: 10.17632/t5sssfgn2v.3.
[26] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: a method for automatic evaluation of machine translation," *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics(ACL)*, pp. 311-318, 2002, doi: 10.3115/1073083.1073135.

BIOGRAPHIES OF AUTHORS

Parag Jayant Mondhe is a Ph.D. Research Scholar at Matoshri College of Engineering and Research Centre, Nashik, India. He is working as an Assistant Professor at K. K. Wagh Institute of Engineering Education and Research, Nashik, India since 2014. He has completed Master of Engineering and Bachelor of Engineering from Savitribai Phule Pune University, India in 2014 and 2012 respectively. His research papers are published in journals and proceedings of international conferences indexed by Scopus. His area of interest includes signal processing, artificial intelligence, and embedded systems. He can be contacted at email: mondheparag@gmail.com.

Dr. Manisha P. Satone is a Professor of Electronics and Telecommunication Engineering at Matoshri College of Engineering and Research Centre, Nashik, India. She was awarded Ph.D. by Savitribai Phule Pune University, India in 2015. She has research and teaching experience of more than 31 years. Her research papers are published in journals indexed by Web of Science and Scopus. She has fetched a research grant and acquired copyrights and patents. She has provided consultancy to many multinational companies. Her area of expertise includes signal processing, artificial intelligence, and embedded systems. She can be contacted at email: mps.eltx@gmail.com.

Dr. Namrata N. Wasatkar is working as an Assistant Professor in the Department of Computer Engineering, Vishwakarma Institute of Information Technology, Pune, India. She was awarded Ph.D. by Savitribai Phule Pune University, India in 2022. She has teaching experience of more than 10 years. Her research papers are published in journals indexed by Scopus. She has fetched a research grant from Savitribai Phule Pune University, India. Her area of expertise includes machine learning, natural language processing, and signal processing. She can be contacted at email: namratakharate1@gmail.com.