Journal of Trends in Computer Science and Smart Technology (ISSN: 2582-4104)
www.irojournals.com/tcsst/

A Comprehensive Study on Sign Language Recognition for Deaf and Dumb people

G. K. Vaidhya[1], C. A. S. Deiva Preetha[2]

Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani, Chennai, Tamil Nadu, India

E-mail: [1]vaidhyagk@gmail.com, [2]deivaprc@srmist.edu.in

Abstract

There are roughly 72 million 'hard of hearing' individuals all over the planet, and more than 80% of them live in developing countries, as indicated in a review by the World Federation for the Deaf. Their lives are hindered by hearing distortions which bar them from showing full interest in the public besides taking pleasure in enjoying identical privileges. Motion based communication is common for the people with hearing and speaking impairments. Communication through signs is a successful choice rather than talking, where the former is replaced by hand flags. One solution to this problem is to study text comprehension tasks for hearing impaired localities using Sign Language Recognition. Gesture-based communication is the most significant and centered approach of communication for deaf and dumb individuals. This paper gives a concise review of different examination works conducted thus far in this field.

Keywords: Sign Language Recognition (SLR), Convolution Neural Network (CNN), Support Vector Machine (SVM), Indian Sign Language (ISL), FFT

1. Introduction

As communication is the method involved with trading thoughts and messages through an assortment of means, gesture-based communication is one of the world's most established and most regular methods for communication [16]. It turns out to be truly challenging for individuals with disabilities to convey, making it very hard for them to interact in their nation's dialects. Bringing many such individuals who have incapacities together, will help add to a more pleasant and more comprehensive world [14]. The gesture-based communication structure can be preferred for cooperation between individuals with hearing hardships to avoid the creation of language boundary. Hand, face, and body gestures

DOI: https://doi.org/10.36548/jtcsst.2022.3.005
Received: 04.08.2022, received in revised form: 24.08.2022, accepted: 09.09.2022, published: 19.09.2022

are used in gesture-based communications to transmit ideas visually [18]. However, gesture-based communications have a lot to offer everyone, including for those who are normal. Deaf or hearing- impaired people frequently collaborate by signing, but this is not their only means of communicating. It is the main way for such people to interface with their messages, and it is basic for others to get a hand on their language.

Computer generated reality frameworks, facial acknowledgment frameworks, and hand-signal acknowledgment frameworks are altogether instances of AI at the activity. However, modern computations like SVM and CNN have significantly improved their usefulness in this field, offering quite a lot of new possibilities. The use of sign language is a customized technique by those who have trouble hearing and tuning in. In India, there is a critical correspondence hole because of the absence of mindfulness about communication through signing [23]. AI and PC vision strategies can be utilized to handle this issue rapidly. To deliver great outcomes, the pre-handling of the dataset of hand movements ought to be extremely exact.

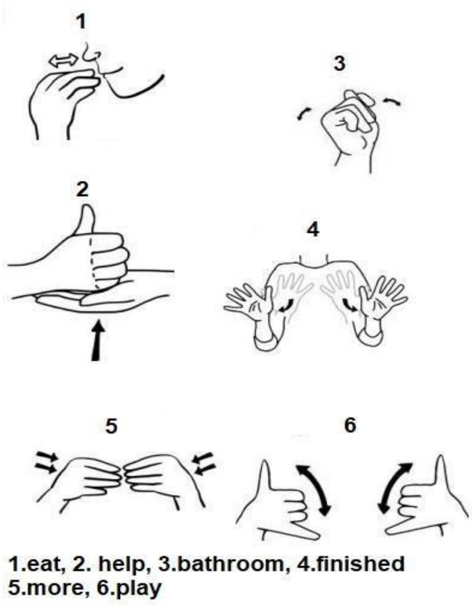

Figure 1. Interpretation of few frequently used gestures

2. Related Work

To create a highly accurate system that is helpful to real-time users, sign language recognition needs effective and reliable data. The data flow for Sign Language Recognition (SLR) is depicted in Fig. 1. Various stages are acquisition of the dataset, preparation of the data, feature extraction, and categorization of the signs.

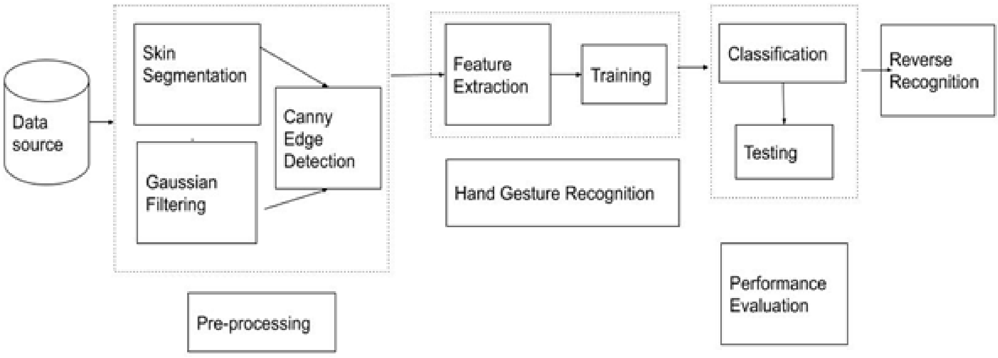

Figure 2. Flow diagram of Sign Language Recognition

2.1 Solutions for SLR

There are many approaches to solve the issue of sign identification, such by using automated systems with video and web cameras for precise gesture classification or by transcribing the sign language used by the deaf into text or voice. Depending on the input data collecting technology, these approaches are primarily split into three key groups:

- Applying various markers to the hands
- Usage of data gloves
- Visual (video or photo) methods of sign recognition
- Infrared sensor (such as those utilized by Microsoft Kinect, Leap Motion, etc.)

The latter two techniques, which are the most popular, enable users to recognize sign language in real time by gathering motions with web cams, leap motion, or other comparable technologies, which is currently available at a reasonable price [4]. Technologies utilizing Artificial Neural Networks (ANN), evolutionary algorithms, hidden Markov models, etc. are employed to recognize sign language in a wide range of theoretical and experimental studies. The selection of a suitable approach for aparticular sign recognition problem requiresa lot of time and effort because there are so many classification and recognition algorithms available. The existing situation indicates a number of issues in this field connected to the collection and processing of input data as well as an ideal recognition method. The part after that offers an analysis of various ANN data preparation techniques for recognizing signs, which are reliable and are often employed in technology for many job domains [5][6].

2.2 Data Pre-Processing

The required input data collection and preprocessing are necessary for each activity, including the recognition of sign language. Most applications use web or video cameras to record their input data [7]. The hardware's low price is the most crucial component. The Myo Gesture Control Armband is one such device, and it costs about 170 British pounds [8]. The data glove technology also has several shortcomings. For instance, less sophisticated gloves don't reveal much about the gesture, and more sophisticated equipment isn't practical for people on a budget. It is also necessary to put on and take off the data gloves each time in order to interpret the motions, which creates additional challenges, particularly for those with disabilities, and renders them inappropriate for installation in public locations.

For real-time sign language recognition, infrared sensor technology, as used in Leap Motion or Microsoft Kinect, currently offers the best prospects. Since there aren't many scholarly publications in this area, this could be a very promising area for future research. The use of video recording technologies for static pictures like national alphabets is frequently debated in literature.

2.2.1 An illustration using Bosnian language recognition

From the article demonstrating Bosnian finger spelling letter identification using ANN [9], an example of data collection and preparation is studied. A database of pictures from three people who had been taught to spell the alphabet with their fingers was used. 90 photos total make up the database, with two-thirds of them being used for system testing and the remaining one-third for training. Several methods were employed to preprocess the data.

An image can be made smaller by converting it from color to grayscale, utilizing histogram equalization to increase contrast, or displaying an image as a spectrum of frequencies using the FFT. Mask-based image segmentation was used for better processing picture binarization to improve object images by removing background, for object recognition border of such an image to only record the most crucial details while minimizing the data.

Grayscale conversion of images is a highly popular operation that enables less data to be handled. In this scenario, it is possible to keep all significant gesture- related information. This example removed all but the brightness of each pixel from an image that was originally 375 by 500 pixels in size.

Histogram equalization increases an image's contrast to get rid of extraneous artifacts that could drastically slow down learning. Each pixel can be classified as either class 1 (foreground) or class 0 using the concept of image binarization (background), removing background noise and allowing the user to concentrate only on pertinent elements. Different techniques that decrease the amount of data and separate key information from irrelevant data can be used to conduct image edge detection. The clever edge detection algorithm was selected. The results of this technique are displayed in Fig. 2, which shows image preprocessing from the original image.

2.2.2 Real-time SLR Using ANN

ANN can be used to recognize sign language in real-time, as well as to capture images and collect video utilizing cameras. At first, the preprocessing comprises of filtering method by removing the noise from the image. The next phase includes the division of input images into 2 groups: Hand movement and Hand shape. The hand shape is identified by each fingertip and its movement direction is identified by the movement by the centre of the palm. Using the Sliding Window technique, the segment size is increased starting with the first set of data points, and the approximation error is calculated for each segment until the maximum error threshold is reached. As soon as all the segment values have been changed, the algorithm creates another segment from i-1 data points.

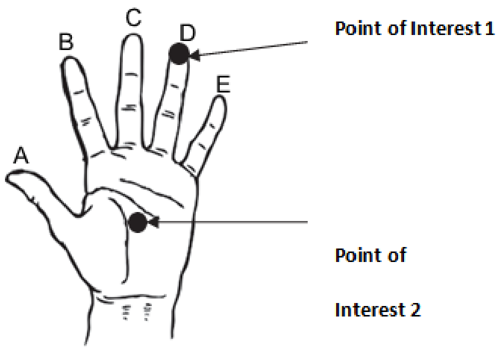

Figure 3. Points of interest in the palm for enhanced gesture identification [12]

The 2 dots specify the points of interest, where the fingertip determines the gesture and the second point indicates the direction of hand movement in sectors. Then the Artificial Neural Network can be used to recognize the gesture.

2.2.3 Segmentation

The segmentation algorithms include splitting and limiting of data. This method is utilized for time series transformation [13]. In SLR, splitting the large pixels into many small segments is doneto increase the performance of the sign recognizing. The author has tried with a few segmentation algorithms in [14]. It incorporated top-down, bottom-up, and sliding window approaches.

2.2.4 Classification

After feature extraction and detection are completed, classification is performed. The two classification techniques are SVM and CNN.

a) Support Vector Machine

A supervised model called the Support Vector Machine (SVM) can tackle classification and regression problems. There are both linear and non-linear challenges. Its feature vectors classify and identify Indian Sign Language. Signs are histograms of visual words that have been provided to it. All images must be used for training, and performance evaluation criteria such as Accuracy, Precision, and Recall, which are used to check the results.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ Score = \frac{2 * (Recall * Precision)}{Recall + Precision}$$

b) Convolutional Neural Network

CNN was influenced by the visual cortex in the human brain. It compares twophotos by roughly identifying the aspects that are comparable in each of the pieces that make up the images, which are referred to as features. Better than any other neural network, CNN can classify theimages.

3. Literature Survey

VibhuGupta [1] proposed a CNN based strategy for unraveling communication through signing and afterward changing it over to message. In the paper, the principal center was around fingerspelling and an extra element of feeling acknowledgment to help the understanding with the third part of communication through signing i.e., non- manual highlights, an ongoing answer for simple translation of gesture-based communication for ordinary individuals as well as deaf and dumb individuals utilizing CNN, thus breaking the language obstruction. The trial results show that this paper accomplished a precision score of 99.8% on testing information.

Bilel Ben Atitallah [2] implemented the CNN Classification technique for hand sign developments, and consuming a low unpredictability with fusing a mechanized purpose in communication, utilizing eight cathodes put on the lower arm, a G-Newton picture multiplication computation, a CNN based finger sign classification and a PC produced hand model for portrayal. The correspondence between the reproduced pictures and ordinary muscle leading to finger images was investigated.

Using an Adam smoothing out specialist and guidance from a dedicated report, a novel CNN classification approach was carried out and improved. In comparison to the outcomes of a SVM and a Softmax classifier, CNN's findings appeared differently. They separately displayed classification accuracy of 95.94 %, 75.61 %, and 62.9 %.

Qazi Mohammad Areeb [3] proposed that the deep learning- based finger motion recognition simulations are created to precisely anticipate the crisis indications of Indian Sign Language (ISL). The dataset utilized covered the recordings for more unique crisis circumstances. A few edges were extricated from the recordings and were taken care of three unique prototypes. Two prototypes were intended for order, while one was an item discovery model applied in the wake of commenting on the edges. The primary model comprised of a three-layered CNN (3D CNN), while the additional involved a pre-prepared VGG-16 and a Repetitive Neural Network (RNN-LSTM) with a long temporary memory conspire. The model depended on YOLO (You Only Look Once) v5, a high-level item location calculation. The forecast correctness of the characterization models were 82% and 98%, separately. Results accomplished a noteworthy mean normal accuracy of 99.6%.

Yangjing Zhou [4] planned a communication through signing acknowledgment model in light of LSTM Neural Network and did the analysis in the open dataset of SLR. It was applied to the improvement of preparing union course of communication for SLR model in light of BiLSTM Neural Network. Hao Zhou [5] introduced the methodology video-based communication for SLR with multi- prompt learning and proposed a Spatial- Temporal Multi-Cue (STMC) organization to address the vision-based succession learning issue. The SMC module figured out how to spatial portray various signals with an independent posture assessment branch. The TMC module modeled worldly adjustments from intra-signal and between prompt viewpoints to investigate the cooperation of numerous signals. A joint streamlining procedure and a sectioned consideration instrument were intended to make the best of multi-sign hotspots for SL acknowledgment and interpretation.

Zaw Hein [6] carried out movement-based gesture-based communication acknowledgment framework that included extraction technique and acknowledgment utilizing AI in view of Myanmar National Sign Language. He proposed skin shading-based upgrade technique and shading-based division strategy for recognizing skin shade of hands, and manual indications of Myanmar Sign Language Recognition System in view of AI.

With the use of machine learning and artificial intelligence, Babita Sonare [7] proposed a fully-fledged Real-time motion- based communication translation structure. The test made use of efficient and superior calculations. By reducing the dataset size restriction, a new RNN-based architecture was employed to improve the precision and execution. After a certain amount of time, the precision remained constant. Through advancements in signing and dynamic pointer motion, the framework also succeedsed admirably in multifarious communication. Ismail HakkiYemenoglu [8] fostered a framework that can decipher gesture-based communication which incorporates letters for individuals who aren't acquainted with communication via gestures.

CNN GoogleNet was employed with move learning strategy. The precision of this communication through sign recognition framework was 91.02%. Mitashi Bansal [9] proposed a hand motion recognition framework utilizing Machine Learning. The dataset of 43000 pictures of letters in order (A-Z) and numbers (0-9), contains 1200 pictures each, were gathered. Then, the pictures were stacked into information model and the model was prepared for sign recognition. PC vision's calculation was utilized to identify signals. Then, classical method was prepared over Machine Learning. Sklearn was utilized for preparing. The model was prepared to take input with accuracy of 96.25%. Varsha M [10] attempted to perceive ISL (Indian Sign Language) signals and convert it into text. At present, a picture recognition model was executed utilizing profound CNN (Inception V3 model) which acknowledges input picture which has gone through a progression of layers and the result was created. It has accomplished an accuracy of 93%.

4. Limitations and Challenges

Most significant prerequisite for gesture-based communication recognition framework is the dataset. Datasets for dialects, for example, American and Arabic gesture-based communication are effectively accessible on the web; however, for different dialects like Indian gesture-based communication, a specific dataset isn't accessible on the web. Therefore, a manual dataset has been created by each of the creators who have contributed. Information increase can be completed on little datasets to make variety of pictures that at last grows the dataset and makes the arrangement model more vigorous. Models depicted in this review give unfortunate outcomes if the dataset incorporates references of benefactors, as the model evaluation produce inaccurate data, and the same issue happens with the shade of the picture. These models additionally deal with issues if they are prepared on shading pictures and the complexion in testing pictures vary from preparing pictures. While working with recordings, the models invest in some opportunity to anticipate sign and the disabled individuals are adjusted with gesture-based communication so their speed can't be coordinated with these current frameworks.

Non-manual methodology could be more compelling in recognition, as it considers looks alongside hand motions, yet it can build the intricacy of execution since variety in look and non-verbal communication can be a lot higher contrasted with variety close by tokens of various individuals. However, sensor-based strategies additionally give more exact outcomes in acknowledgment, and they experience in parts of compactness and reasonableness.

5. Conclusion

In this paper, a study on communication via gesture recognition is introduced and different strategies have been read up and examined for something similar. In recognition process, segmentation has a critical impact wherein skin spot is isolated from the picture which typically influences the accuracy of recognition. Other than segmentation, classification additionally relies upon the feature extraction procedures which performs dimensionality decrease and reduces the calculation cost. Investigation of different grouping procedures presumes that CNN, Inception model, and LSTM perform better when compared to the conventional classifiers like KNN and SVM.

References

[1] V. Gupta, M. Jain and G. Aggarwal, "Sign Language to Text for Deaf and Dumb," 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2022, pp. 384-389, doi: 10.1109/Confluence52989.2022.9734196.

[2] B. Ben Atitallah et al., "Hand Sign Recognition System Based on EIT Imaging and Robust CNN Classification," in IEEE Sensors Journal, vol. 22, no. 2, pp. 1729-1737, 15 Jan.15, 2022, doi: 10.1109/JSEN.2021.3130982

[3] Q. M. Areeb, Maryam, M. Nadeem, R. Alroobaea and F. Anwer, "Helping Hearing-Impaired in Emergency Situations: A Deep Learning-Based Approach," in IEEE Access, vol. 10, pp. 8502-8517, 2022, doi: 10.1109/ACCESS.2022.3142918.

[4] Y. Zhou, C. Ji and L. Cao, "Research on Optimizer Algorithm of Sign Language Recognition Model," 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA), 2022, pp. 102-106, doi: 10.1109/ICPECA53709.2022.9719010.

[5] H. Zhou, W. Zhou, Y. Zhou and H. Li, "Spatial- Temporal Multi-Cue Network for Sign Language Recognition and Translation," in IEEE Transactions on Multimedia, vol. 24, pp. 768- 779, 2022, doi: 10.1109/TMM.2021.3059098

[6] Z. Hein, T. P. Htoo, B. Aye, S. M. Htet and K. Z. Ye, "Leap Motion based Myanmar Sign Language Recognition using Machine Learning," 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), 2021, pp. 2304-2310, doi: 10.1109/ElConRus51938.2021.9396496.

[7] B. Sonare, A. Padgal, Y. Gaikwad and A. Patil, "Video-Based Sign Language Translation System Using Machine Learning," 2021 2nd International Conference for Emerging Technology (INCET), 2021, pp. 1-4, doi: 10.1109/INCET51464.2021.9456176.

[8] I. H. Yemenoglu, A. F. M. S. Shah and H. Ilhan, "Deep Convolutional Neural Networks- Based Sign Language Recognition System," 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2021, pp. 0573-0576, doi: 10.1109/IEMCON53756.2021.9623068.

[9] M. Bansal and S. Gupta, "Detection and Recognition of Hand Gestures for Indian Sign Language Recognition System," 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC), 2021, pp. 136- 140, doi: 10.1109/ISPCC53510.2021.9609448.

[10] M. Varsha and C. S. Nair, "Indian Sign Language Gesture Recognition Using Deep Convolutional Neural Network," 2021 8th International Conference on Smart Computing and Communications (ICSCC), 2021, pp. 193- 197, doi: 10.1109/ICSCC51209.2021.9528246.

[11] B. Joksimoski et al., "Technological solutions for sign language recognition: a scoping review of research trends, challenges, and opportunities," in IEEE Access, doi: 10.1109/ACCESS.2022.3161440.

[12] X. Han, F. Lu, J. Yin, G. Tian and J. Liu, "Sign Language Recognition Based on R(2+1)D With Spatial–Temporal–Channel Attention," in IEEE Transactions on Human-Machine Systems, doi: 10.1109/THMS.2022.3144000.

[13] A. Alqahtani et al., "Improving the Virtual Educational Platforms for the Deaf and Dumb under the Covid-19 Pandemic Circumstances," 2022 2nd International Conference on Computing and Information Technology (ICCIT), 2022, pp. 191-196, doi: 10.1109/ICCIT52419.2022.9711613.

[14] Y. C. Bilge, R. G. Cinbis and N. Ikizler-Cinbis, "Towards Zero-shot Sign Language Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, doi: 10.1109/TPAMI.2022.3143074.

[15] O. MercanogluSincan and H. Y. Keles, "Using Motion History Images With 3D Convolutional Networks in Isolated Sign Language Recognition," in IEEE Access, vol. 10, pp. 18608-18618, 2022, doi: 10.1109/ACCESS.2022.3151362.

[16] M. Al-Qurishi, T. Khalid and R. Souissi, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," in IEEE Access, vol. 9, pp. 126917- 126951, 2021, doi: 10.1109/ACCESS.2021.3110912.

[17] R. Singh and M. Jangid, "Indian Sign language Recognition Using Color Space Model and Thresholding," 2021 Asian Conference on Innovation in Technology (ASIANCON), 2021, pp. 1-4, doi: 10.1109/ASIANCON51346.2021.9544615.

[18] C. Chu, Q. Xiao, J. Xiao and C. Gao, "Sign Language Action Recognition System Based on Deep Learning," 2021 5th International Conference on Automation, Control and Robots (ICACR), 2021, pp. 24-28, doi: 10.1109/ICACR53472.2021.9605168.

[19] W. Li, H. Pu and R. Wang, "Sign Language Recognition Based on Computer Vision," 2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 2021, pp. 919-922, doi: 10.1109/ICAICA52286.2021.9498024.

[20] S. Tornay, M. Razavi and M. Magimai.-Doss, "Towards Multilingual Sign Language Recognition," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 6309-6313, doi: 10.1109/ICASSP40776.2020.9054631.

[21] Y. Zhang and L. Cao, "A Survey on Neural Machine Translation Applied to Sign Language Generation," 2021 3rd International Conference on Applied Machine Learning (ICAML), 2021, pp. 413-417, doi:10.1109/ICAML54311.2021.00093.

[22] H. N. Saha, S. Tapadar, S. Ray, S. K. Chatterjee and S. Saha, "A Machine Learning Based Approach for Hand Gesture Recognition using Distinctive Feature Extraction,"

2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), 2018, pp. 91-98, doi: 10.1109/CCWC.2018.8301631.

[23] Suharjito, H. Gunawan, N. Thiracitta and A. Nugroho, "Sign Language Recognition Using Modified Convolutional Neural Network Model," 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), 2018, pp. 1-5, doi: 10.1109/INAPR.2018.8627014.

[24] M. Xie and X. Ma, "End-to-End Residual Neural Network with Data Augmentation for Sign Language Recognition," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2019, pp. 1629-1633, doi: 10.1109/IAEAC47372.2019.8998073.

[25] B. Gupta, P. Shukla and A. Mittal, "K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion," 2016 International Conference on Computer Communication and Informatics (ICCCI), 2016, pp. 1-5, doi: 10.1109/ICCCI.2016.7479951.

Author’s biography

G. K. Vaidhya is a Research Scholar in the Department of Computing Technologies, SRM Institute of Science and Technology, Vadapalani, India. She earned her Bachelor of Engineering in Computer Science and Engineering from Shanmuganathan Engineering College of Anna University in Pudukkottai, India, in 2008, and her Master of Engineering in Computer Science and Engineering from Periyar Maniammai University in Thanjavur, India, in 2011. She has ten years of experience in Teaching. Image processing and machine learning are two of her main interests.

C. A. S. Deiva Preetha is working as an Assistant Professor (Senior Grade) in the Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani, India. She received her Bachelor of Engineering in Computer Science and Engineering from Bharatiyar University, Coimbatore, India in 1997 and Master of Science from BITS Pilani, Rajasthan, India in 2003. She was awarded Ph.D. in the Faculty of Information and Technology from SRM Institute of Science and Technology, Kattankulathur, Chennai in 2020. She has 3 years of teaching and 14 years of industry experience. Her area of research specialization is Software Engineering.