Original Article

Real time sign language detection system using deep learning techniques

N. Padmaja[1], B.Nikhil Sai Raja[2], B.Pavan Kumar[3]
[1]Dept of Computer Science and Engineering, Assistant Professor, SPMVV, Andhra Pradesh, India
[2]Year M.Tech(QR OR),Indian Statistical Institute, Kolkata
[3]III Year B.Tech (MECH), NIT, Durgapur, West Bengal
Email: gowripadma99@gmail.com
DOI: 10.47750/pnr.2022.13.S01.126

Abstract

Humans require communication in order to survive. It is a fundamental and effective method for communicating thoughts, feelings, and points of view. A greater number of babies are being born with hearing abnormalities, which puts them at a communication disadvantage with the rest of the world, according to data on physically challenged children during the last decade. People who are deaf or hard of hearing typically use sign languages to communicate. Hand gestures are used by Deaf or Mute persons to communicate; as a result, non-Deaf people have a hard time understanding their messages. Systems that can detect different signs and provide information to common people are thus necessary. To address this issue, we have developed an automated sign language detection system using deep learning, which helps deaf/mute people can communicate with normal people.

Keywords: deaf; hearing problem; communication; normal people; survey; vision based; deep learning; Faster RCNN; ResNet50.

1. INTRODUCTION

In order to function, humans need to be able to communicate. It is a basic and productive approach to express ideas, emotions, and viewpoints. However, this skill is absent from a sizeable portion of the global population [1]. Many people experience either hearing loss, speech impairment, or both. Hearing loss refers to a partial or total loss of hearing in one or both ears. On the other side, mute is a disability that prevents speech and renders those who have it unable to talk. Hearing mutism, also known as deaf-mutism, is a linguistic deficit that can develop in people who become deaf-mute when still children. These conditions are among the most prevalent impairments in the world. According to statistics on physically challenged children during the past ten years, more newborns are being born with hearing defects, which puts them at a communication disadvantage with the rest of the world [2]. According to the WHO's most recent data, 466 million individuals have hearing loss in 2019.

It may be quite difficult to communicate with those who have hearing impairments. Worldwide, deaf and hard of hearing persons mostly communicate using sign languages [3]. The social connection and communication gap between them and the able-bodied persons can be filled up most powerfully and effectively through sign language. People who are Deaf or Mute utilize hand gestures to communicate; as a result, regular people have difficulty understanding the signals they make. Therefore, there is a need for systems that can identify various signs and communicate information to regular people.

1.1 Sign Language

As a visual language, sign language is used. Basically, it is made up of these 3 primary parts:

1. Word - based association, which entails hand movements which express the meanings of words.
2. Finger spelling: Invoke out words character by character.
3. Non-manual features: Body positions, mouth, tongue and facial expressions.

Over 70 million people use more than 300 sign languages globally, according to the World Federation of the Deaf [5].

1.2 Current methods

The following techniques are primarily used to identify sign gestures:

1. The glove-based technique, which requires the signer to wear a hardware glove whereas the motions of their hands are being recorded.

2. Static and dynamic recognition techniques based on vision. While dynamic involves the live capturing of the movements in real-time, statics deals with the recognition of static motions (i.e., 2D pictures). Using the camera to record movement is required for this.

Despite being over 90% accurate, the glove-based approach appears a little unpleasant for usage in real-world situations. In this paper we proposed a vision based approach for word level and finger spelling sign language.

Despite the fact that sign language is crucial for deaf-mute persons to communicate with others and with themselves, normal people nevertheless pay it little mind. Unless we have relatives who are deaf-mute, we as normal people often overlook the value of sign language. Using a sign language interpreter's services is one way to communicate with deaf-mute persons [4]. However, hiring a sign language interpreter might be expensive. For the deaf-mute and normal individuals to converse regularly, a low-cost solution is necessary.

Therefore, scientists are working to develop a means of communication for deaf-mute persons so that they may interact with hearing people. The Sign Language Recognition System is the innovation in this. The technology seeks to understand sign language and translate it either orally or in written form into the native tongue. However, developing this technology was incredibly expensive and challenging to implement for everyday usage. Early studies have shown that employing data gloves to recognize sign language is successful. Nevertheless, it is challenging to market due to the high cost of the gloves and wearing character. Researchers then attempt to create a pure vision Sign Language Recognition System in light of this knowledge. However, it also has drawbacks, particularly when trying to properly monitor hand motions.

From picture capture to classification, issues in developing sign language recognition might arise. The ideal technique for acquiring images is currently being investigated by scientists. The challenges of image pre-processing are presented by image collection using a camera. Using an active sensor device, meanwhile, can be expensive. Researchers' use of classification systems has certain downsides as well. Researchers are unable to choose the optimum recognition method since there are so many options. By concentrating on one approach, it tends to prevent testing of other methods that could be more appropriate for sign language recognition. Researchers seldom fully develop one strategy to its potential by trying out other approaches. The purpose of this study is to explain the Sign Language Recognition System that researchers employ and also develop a sign language detection which helps to have a better communication with mute people.

2. Literature Survey

The field of touch recognition research is highly competitive. It could entail a range of sensory impressions comparable to the usage of sensory equipment[40]. However, in real life, employing hardware is both expensive and difficult. Thus, researchers are searching for the most accurate detection utilizing computer-aided detection approaches. The effectiveness of learning sign language based on various approaches and viewpoints is compared with [6]. Because they may extract characteristics from a website's raw code, in-depth learning algorithms are both simple to apply and advantageous. Performance of the system is affected by the depth of Convolutional Neural Networks and the mix of CNN layers.

In terms of computer science, SL is one of the most active research fields and has put out a lot of fresh ideas. Several techniques, including Artificial Neural Networks [8], Hidden Markov Models [9], and Advanced Learning utilizing 3D CNN [9], have been utilized by researchers to identify SL. CNN is a type of artificial neural network that can identify characteristics in a picture [11][38]. ResNet, GoogleNet, VGGNet, AlexNet, and LeNet [12], which won the ImageNet Large Scale Vision Recognition Challenge starting in 2010, are CNN's most well-liked formats. Many applications, including object identification, group reconstruction, numerical recognition, SL character recognition, video tracking, and many more, have been developed using the best outcomes from these techniques[37][39].

To create better feature maps for digitization, morphological pattern recognition exercises employing the opposite harmonic definition were paired with a conversion layer [15][34]. A superior outcome was obtained using the morphological CNN evaluated on the MNIST digital database[28][29][44].

A suitable technique for an ASL recognition system was put forth [7][33][45][46]. They have previously assessed the symbols' effects, assembled different regional structures in the processed photos, and depending on the stated features, the signed touches were converted into text. This technique can identify 10 numerals and 24 ASL characters. A comparison between the CNN-based method and the contour-support vector (SVM) machine was conducted [13][31]. They have employed SLD, ASL, and ASL-FS, three separate standard information sites, in a range of situations including rotation and backdrop measurement[35][36][47][49]. The concert was purchased and coated as a convex shell using a cement method that was focused on width and convex angle. The touch was then separated using SVM. The accuracy and precision rates for the suggested comparison were respectively 69 percent and 98.31%. The MNIST and CIFAR-10 databases employed the CNN algorithm [14][48]. In these picture databases, they examined the algorithm's capacity to recognize and detect symbols. Table 1 displays related research on sensor-based sign language recognition and machine learning[30][32][43].

Table 1. Related research works on sensor based and machine learning based sign language recognition

Ref	Dataset	Classifier	Feature Extraction	Remarks
[17]	27 lexical items	Linear SVM	Tracking 2D and 3D structure	Lower accuracy
[27]	Persian 46 words and 20 sentences	KNN-DTW, HMM	Skin segmentation, Head/hands tracking	93.73 percent accuracy
[24]	30 Arabic signs	Min. Dist, KNN, SVM	LDA and SIFT	90 percent accuracy
[26]	239 signs	3D trajectory match algorithm	3D hand track	96.32 percent accuracy
[20]	19 signs	HMM	Six Gaussians for BP and HS from joint angles	76.2 percent accuracy
[19]	24 hand shapes	Random Forest	Depth and Appearance images	75 percent accuracy
[23]	3 different datasets	DP Variant Enhanced Level Building	Motion and skin segmentation	70 percent accuracy
[18]	48 signs	CRF threshold	Hand and face detection, tracking	93.5 percent accuracy
[21]	25 sentences 183 signs	Multi HMM	Frame difference Positional coordinate measure	75.6 percent accuracy
[22]	120 Dutch signs	SDTW+Q-DFFM SDTW+CDFD	Skin segmentation Blob tracking	92.3 percent accuracy achieved by SDTW+CDFD
[16]	5113 Chinese signs	Transition movement	2 Datagloves and Pohelmus	91.9 percent accuracy
[25]	104 American SL signs 201 sentences	PCA HMM	Head/Hand position tracking Image Scaling	Accuracy is low

3. Methodology

This paper presents the methodology for sign language detection in real time images. Deep detector: Faster R-CNN (Faster Region-based Convolutional neural network) is fed 200 photos taken with different resolutions using a mobile phone camera at the front and side views, and it combines with "deep feature extractor: Residual Network-50 (ResNet-50)" to create a more suitable deep learning architecture to detect signs[41][43].

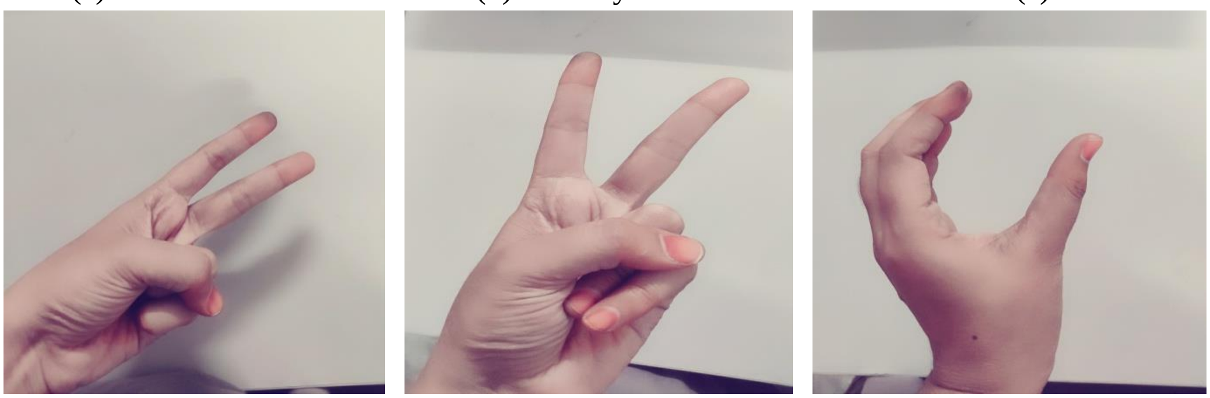
3.1 System Overview

Finding several signs is the aim of this study. Figure 1 shows the system from a high perspective. Below, each stage of the proposed system is thoroughly explained.

3.2 Data Collection and Data annotation

Sign dataset has images of various signs i.e. of front view and side view. Sign dataset contains 200 images total and has sign data like Letter V, L, U and one Word I Love you for each sign we captured 50 samples of various resolutions images (48MP, 12MP). Figure 2 represents different sign samples considered for the proposed system. Annotated all the images in the dataset manually using label Img tool by drawing the bounding where the sign is present and specifies the label as the class type to which sign belongs to.

(a) I Love You Word (b) I Love you Word (c) Letter L



(d) Letter V (side) (e) Letter V (front) (f) Letter U

Figure 2: Sample images in the sign dataset

3.3 Faster-RCNN+ResNet50 Sign Language Detection

The goal is to detect and classify the type of sign i.e Letter V, L, U and I Love you word. The bounding boxes in which the sign must be accurately established for the system to produce reliable results. The Region Proposal Network accelerates R-generation CNN's of crucial ROIs for sign detection (RPN).

A Faster RCNN approach for sign identification in a picture follows these steps:

1. Send an image to the ConvNet in order to provide RPN with feature maps for the picture.

2. RPN slides windows on acquired feature maps at each window to construct k fixed - sized anchor boxes with different shapes and sizes in a sign picture, and predicts the chance that an anchor is a sign as well as the bounding-box regressor that will best fit the anchor for sign class.

3. After getting various kinds and sizes of bounding boxes and clipping each proposal so that every proposal contains signs, ROI gathers fixed-size feature mappings to all the anchors.

4. The gathered fixed size feature mappings are sent to a fully connected layer that has a linear regression layer and a softmax layer. The sign is categorized and estimated bounding-boxes for the detected signs at the end.

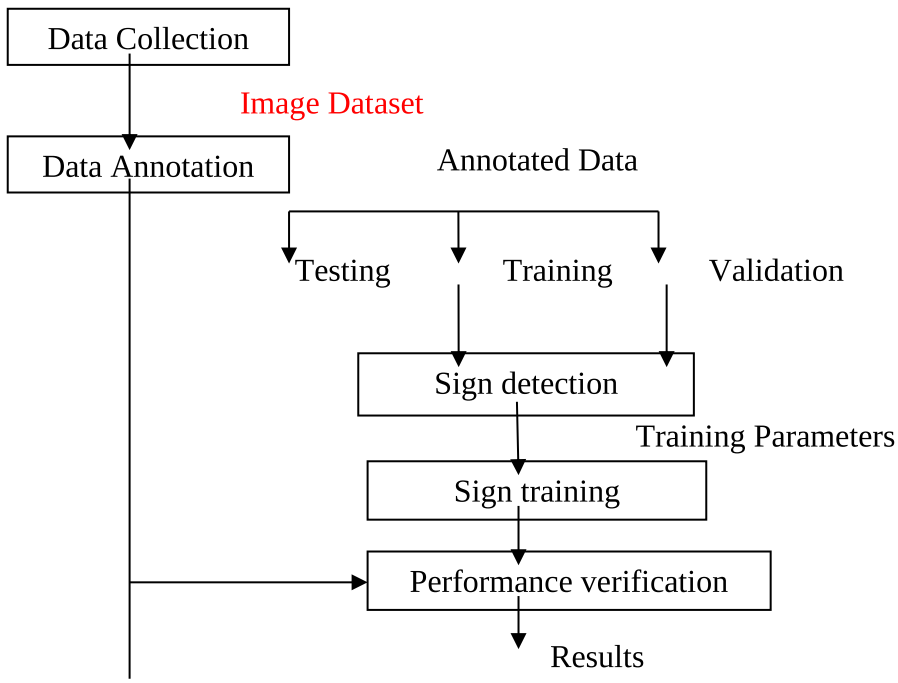


Figure 1. High-level picture of proposed system for sign language detection

4. Experimental results

The suggested system employs ResNet-50 and Faster R-CNN to detect signs. IoU and AP [11] are used to assess the performance of the proposed system.

$$\mathrm{IoU}(A, B) = \left| \frac{A \cap B}{A \cup B} \right| \qquad (1)$$

Here,

IoU : Intersection over Union

A : Annotated ground bounding box

B : Bounding box predicted by the system

$$AP = \frac{1}{11}\sum_{r\in\{0,\ 0.1,\dots,1\}} \max_{\check{r}:\,\check{r}\geq r} p(\check{r}) \qquad (2)$$

Here,

AP : Average Precision

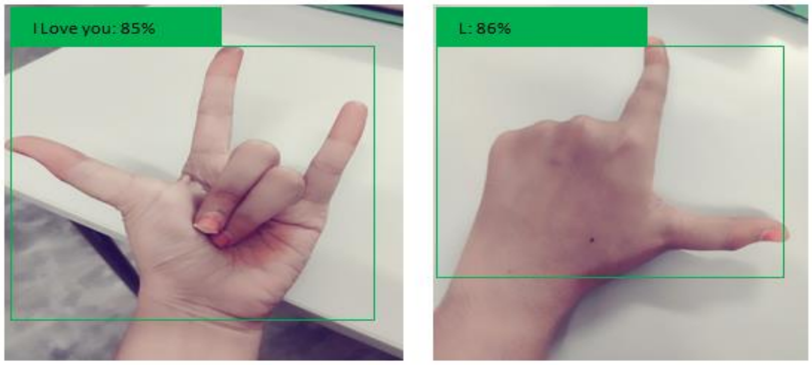
p(ř) : Precision at recall

When the calculated IoU exceeds the threshold value of 0.5, the expected result is classified as either a true positive (TP) or a false positive (FP). Mean Average Precision (mAP) is the AP calculated for different signs. Table 2 shows the quantitative results achieved by the proposed system. Accuracy of our proposed sign language detection system is 86 percent and the qualitative results predicted by the proposed system is shown in figure 3.

Table 2. Faster RCNN+ResNet50 quantitative results achieved for different sign classes

Class	**Fatser RCNN+ResNet50**
Letter V	85%
Letter L	86%
Letter U	87%

I Love You word	88%
Total mean AP	**86%**



(a) I Love You Word (front) (b) Letter L

(c) Letter V (side) (d) Letter V (front) (e) Letter U

Figure 3. Qualitative results achieved by the system

5. Conclusion

People who are deaf or hard of hearing typically use sign languages to communicate. Hand gestures are used by Deaf or Mute persons to communicate; as a result, non-Deaf people have a hard time understanding their messages. The main methods for identifying sign gestures include glove-based recognition approaches as well as static and dynamic vision-based techniques. The glove-based method looks a tad uncomfortable to use in real-world conditions, while being over 90% accurate. We suggested a vision-based method for word-level and finger-spelling sign language in this research. A total of 200 photos make up the sign dataset, which also includes sign data such as the letters V, L, and U and the word I love you for each sign (48MP, 12MP).

REFERENCES

[1] El-Din, S. A. E., & El-Ghany, M. A. A. (2020). Sign language interpreter system: An alternative system for machine learning. In Proceedings of the 2nd novel intel- ligent and leading emerging sciences conference, NILES 2020, Ml (pp. 332–337). https://doi.org/10.1109/NILES50944.2020.9257958 .

[2] Hasan, Md Mehedi, Srizon, A. Y., Sayeed, A., & Hasan, M. A. M (2020). Classification of sign language characters by applying a deep convolutional neural network. In Proceedings of the 2nd international conference on advanced information and communication technology, ICAICT (pp. 434–438). 2020, November. https://doi. org/10.1109/ICAICT51780.2020.9333456 .

[3] Izzah, A., & Suciati, N. (2014). Translation of sign language using generic fourier descriptor and nearest neighbour. International Journal on Cybernetics and Infor- matics, 3 (1), 31–41. https://doi.org/10.5121/ijci.2014.3104 .

[4] Kudrinko, K., Flavin, E., Zhu, X., & Li, Q. (2021). Wearable sensor-based sign language recognition: A comprehensive review. IEEE Reviews in Biomedical Engineering, 14 , 82–97. https://doi.org/10.1109/RBME.2020.3019769 .

[5] Rastgoo, R., Kiani, K., & Escalera, S. (2021a). Hand pose aware multimodal isolated sign language recognition. Multimedia Tools and Applications, 80 (1), 127–163. https://doi.org/10.1007/s11042- 020- 09700- 0 .

[6] Sruthi, C. J., and A. Lijiya. "Signal: An in-depth learning program based on the Indian Sign Language Recognition Program." 2019 International Conference on Communication and Signal Processing (ICCSP), pp. 0596-0600. IEEE, 2019.
[7] Shivashankara S., Srinath S. American Sign Language Awareness Program: The Right WayInt. J. Picture graph. Signal Process., 10 (2018)
[8] Kishore P., Prasad M., Prasad C.R., Rahul R.4 sign language camera model using elliptical fourier adjectives and ANN2015 International Conference on Signal Processing and Communication Communications Engineering Systems (2015), pp. 34-38
[9] Zhang J., Zhou W., Xie C., Pu J., Li H.Chinese Sign Language Recognition with Variable HMM2016 IEEE International Conference on Multimedia and Expo (ICME) (2016), pp. 1-6
[10] Camgoz N.C., Hadfield S., Koller O., Bowden R.Convolutional 3d neural networks are used to obtain continuous user continuous touch recognition2016 23rd International Conference on Pattern Recognition (ICPR) (2016), pp. 49-54
[11] Camgoz N.C., Hadfield S., Koller O., Bowden R.Convolutional 3d neural networks are used to obtain continuous user continuous touch recognition2016 23rd International Conference on Pattern Recognition (ICPR) (2016), pp. 49-54
[12] Khan A., Sohail A., Zahoora U., Qureshi A.S.A survey of the latest structures for convolutional n neural networks(2019)
[13] Kalbhor S.R., Deshpande A.M.Digital recognition using machine learning and Convolutional Neural NetworkThe Second 2018 International Conference on Electronics Styles and Informatics (ICOEI) (2018), pages 604-609
[14] hauhan R., Ghanshala K.K., Joshi R.Convolutional Neural Network (CNN) Image Acquisition and Visibility2018 First International Conference on Computer Security and Communication (ICSCCC) (2018), pages 278-282
[15] Rao, G. Anantha, K. Syamala, P. V. V. Kishore, and A. S. C. S. Sastry."Deep convolutional networks to recognize sign language."2018 Conference on Symbol Processing and Communication EngineeringSystems (SPACES), pages 194-197. IEEE, 2018.
[16] Fang, Gaolin, Wen Gao, and Debin Zhao. "Large-vocabulary continuous sign language recognition based on transition-movement models." IEEE transactions on systems, man, and cybernetics-part a: systems and humans 37.1 (2006): 1-9.
[17] Zhang, Chenyang, Yingli Tian, and Matt Huenerfauth. "Multi-modality American sign language recognition." 2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016.
[18] Yang, Hee-Deok, Stan Sclaroff, and Seong-Whan Lee. "Sign language spotting with a threshold model based on conditional random fields." IEEE transactions on pattern analysis and machine intelligence 31.7 (2008): 1264-1277.
[19] Pugeault, Nicolas, and Richard Bowden. "Spelling it out: Real-time ASL fingerspelling recognition." 2011 IEEE International conference on computer vision workshops (ICCV workshops). IEEE, 2011.
[20] Zafrulla, Zahoor, et al. "American sign language recognition with the kinect." Proceedings of the 13th international conference on multimodal interfaces. 2011.
[21] Maebatake, Masaru, et al. "Sign language recognition based on position and movement using multi-stream HMM." 2008 Second International Symposium on Universal Communication. IEEE, 2008.
[22] Lichtenauer, Jeroen F., Emile A. Hendriks, and Marcel JT Reinders. "Sign language recognition by combining statistical DTW and independent classification." IEEE transactions on pattern analysis and machine intelligence 30.11 (2008): 2040-2046.
[23] Yang, Ruiduo, Sudeep Sarkar, and Barbara Loeding. "Handling movement epenthesis and hand segmentation ambiguities in continuous sign language recognition using nested dynamic programming." IEEE transactions on pattern analysis and machine intelligence 32.3 (2009): 462-477.
[24] Tharwat, Alaa, et al. "Sift-based arabic sign language recognition system." Afro-european conference for industrial advancement. Springer, Cham, 2015.
[25] Rybach, David, et al. "Speech recognition techniques for a sign language recognition system." In Interspeech 2007-Eurospeech. 2007.
[26] Chai, Xiujuan, et al. "Sign language recognition and translation with kinect." IEEE Conf. on AFGR. Vol. 655. 2013.
[27] Zadghorban, Masoud, and Manoochehr Nahvi. "An algorithm on sign words extraction and recognition of continuous Persian sign language based on motion and shape features of hands." Pattern Analysis and Applications 21.2 (2018): 323-335.
[28] Sunil Kumar, M., and A. Rama Mohan Reddy. "An Efficient Approach for Evolution of Functional Requirements to Improve the Quality of Software Architecture." In Artificial Intelligence and Evolutionary Computations in Engineering Systems, pp. 775-792. Springer, New Delhi, 2016.
[29]Kumar, M. Sunil, et al. "Automated Extraction of Non-Functional Requirements From Text Files: A Supervised Learning Approach." Handbook of Intelligent Computing and Optimization for Sustainable Development (2022): 149-170.
[30] Swamy, B. M., & Kumar, M. S. (2021). A comparative study on a privacy protection in personalized web search. Materials Today: Proceedings.
[31] V Anantha Natarajan, V Tamizhazhagan, Naresh Tangudu, M Sunil Kumar,"Analysis of Groundwater Level Fluctuations and its Association with Rainfall Using Statistical Methods",journal of algebraic Statistics, Vol13. no.3, 2022.
[32] V AnanthaNatarajan,M Sunil Kumar,V Tamizhazhagan, "Forecasting of Wind Power using LSTM Recurrent Neural Network", Journal of Green Engineering, Volume-10, Issue-11, November 2020.
[33] Natarajan, V. Anantha, M. Sunil Kumar, Rizwan Patan, Suresh Kallam, and Mohamed Yasin Noor Mohamed. "Segmentation of nuclei in histopathology images using fully convolutional deep neural architecture." In 2020 International Conference on computing and information technology (ICCIT-1441), pp. 1-7. IEEE, 2020.
[34] Natarajan VA, Babitha MM, Kumar MS. Detection of disease in tomato plant using Deep Learning Techniques. International Journal of Modern Agriculture. 2020 Dec 30;9(4):525-40.
[35] Gampala, Veerraju, M. Sunil Kumar, C. Sushama, and E. Fantin Irudaya Raj. "Deep learning based image processing approaches for image deblurring." Materials Today: Proceedings (2020).
[36] Ganesh, D., Kumar, M. S., Reddy, M. P. V., Kavitha, S., & Murthy, D. S. Implementation of AI Pop Bots and its allied Applications for Designing Efficient Curriculum in Early Childhood Education. International Journal of Early Childhood, 14(03), 2022.
[37] Sushama, C., Arulprakash, P., Kumar, M. S., Ganesh, D., & Sujatha, K. The Future of Education: Artificial Intelligence based Remote Learning. International Journal of Early Childhood, 14(03), 2022.
[38] Balaji, K., P. Sai Kiran, and M. Sunil Kumar. "Power aware virtual machine placement in IaaS cloud using discrete firefly algorithm." Applied Nanoscience (2022): 1-9.
[39] Prakash, M. S. S., Bandlamudi, M., & Radhakrishnan, R. Educating and communicating with deaf learner's using CNN based Sign Language Prediction System", International Journal of Early Childhood Special Education, Vol 14, Issue 02, 2022.
[40] Prasad, P. Yogendra, Dumpa Prasad, D. Naga Malleswari, Monali N. Shetty, and Neha Gupta. "Implementation of Machine Learning Based Google Teachable Machine in Early Childhood Education." International Journal of Early Childhood 14, no. 03: 2022.
[41] Peneti, Subhashini, et al. "BDN-GWMNN: internet of things (IoT) enabled secure smart city applications." Wireless Personal Communications 119.3 (2021): 2469-2485.
[42] Davanam, G., Pavan Kumar, T., & Sunil Kumar, M. (2021). Novel Defense Framework for Cross-layer Attacks in Cognitive Radio Networks. In International Conference on Intelligent and Smart Computing in Data Analytics (pp. 23-33). Springer, Singapore.
[43] Ganesh, Davanam, Thummala Pavan Kumar, and Malchi Sunil Kumar. "Optimised Levenshtein centroid cross-layer defence for multi-hop cognitive

radio networks." IET Communications 15.2 (2021): 245-256.
[44] Davanam, Ganesh, T. Pavan Kumar, and M. Sunil Kumar. "Efficient energy management for reducing cross layer attacks in cognitive radio networks." Journal of Green Engineering 11 (2021): 1412-1426.
[45] Malchi, Sunil Kumar, et al. "A trust-based fuzzy neural network for smart data fusion in internet of things." Computers & Electrical Engineering 89 (2021): 106901.
[46] Sreedhar, B., Manjunath Swamy BE, and M. Sunil Kumar. "A comparative study of melanoma skin cancer detection in traditional and current image processing techniques." 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). IEEE, 2020.
[47] Balaji, K., P. Sai Kiran, and M. Sunil Kumar. "Resource aware virtual machine placement in IaaS cloud using bio-inspired firefly algorithm." Journal of Green Engineering 10 (2020): 9315-9327.
[48] Kumar, M. Sunil, and K. Jyothi Prakash. "Internet of things: IETF protocols, algorithms and applications." Int. J. Innov. Technol. Explor. Eng 8.11 (2019): 2853-2857.
[49] Harika, A., Sunil Kumar, M., Anantha Natarajan, V., & Kallam, S. (2021). Business process reengineering: issues and challenges. In Proceedings of Second International Conference on Smart Energy and Communication (pp. 363-382). Springer, Singapore.