COMPARATIVE ANALYSIS OF VARIOUS ACTIVATION FUNCTIONS IN CONVOLUTIONAL NEURAL NETWORK FOR SIGN LANGUAGE RECOGNITION

Sornam M[1]*, Panneer Selvam E[2]

Abstact-
Communication is the act of sharing and receiving information where the modality of spoken language is oral-auditory and visual-gesture for sign language that makes huge difference in interpretation of sign languages. In most of the cases convolutional neural network model is used to recognize the sign language where the activation function is the fundamental component of the hierarchical structure of the CNN model due to its nonlinear properties. On the basis of the Keras framework, seven common activation functions namely, ReLu, LeakyReLu, PReLu, ReLu6, SELU, Swish, HardSwish, and Mish have been analyzed and evaluated in sign language recognition tasks. This work evaluates the performance of Convolutional Neural Network (CNN) with different activation functions and the experimental results shows that CNN based on mish activation function has a phenomenal improvement in performance than the other activation functions.

Keywords: Activation Function, Convolutional Neural Network, Sign Language, Visual Gesture.

[1]*Department of Computer Science, University of Madras, Chennai,INDIA,
E-mail:- Madasamy.sornam@gmail.com
[2]Department of Computer Applications, DDGD Vaishnav College, Chennai, INDIA,
E-mail:- panneerselvame@dgvaishnavcollege.edu.in

***Correspondence Author: -** Sornam M
*Department of Computer Science, University of Madras, Chennai,INDIA,
E-mail:- Madasamy.sornam@gmail.com

DOI: - 10.48047/ecb/2023.12.si5a.0305

I. Introduction

Sign language is a non-verbal visual communication language developed for hearing and speech impaired people around the globe. There is no conclusive evidence regarding the origin of sign language, Plato from ancient Greece first recorded the language in the fifth century BC. According to the World Federation of the Deaf, there are more than 70 million deaf people around worldwide and more 300 sign languages in the globe, Like spoken language, sign languages also distinct from one another. The goal of sign language recognition (SLR) is to create an aid that instantly translates an input sign into the equivalent text or speech. The communication gap between the hearing impaired and the rest of society can be closed with the help of SLR systems. As a result, such systems provide a new way for applications based on human-computer interaction (HCI). Numerous SLR systems have been created by researchers for primary sign languages, but there are only few recognizable work has been done for ISL interpretation.

The deaf community in India uses Indian Sign Language (ISL). However, ISL is not yet recognized as an official language but the process is on. Most forms of Indian sign language are inherited from British sign language, which uses two hands where ASL which is taught in deaf schools in India is taught using only one hand. In India, study of Indian Sign Language first started in 1978. ISL was only used in short-term courses because there was no established standard format for ISL. Indian Sign Language (ISL) uses both single-handed and double-handed gestures, as well as a variety of signs for the same alphabet depending on the region of India. It makes difficult to implement a sing recognition model for ISL. Furthermore, there is a shortage of standard dataset. These factors become challenging for the researchers.

ISL became standardized only in 2003. On March 23, 2018, at the India International Center in New Delhi, Indian Sign Language Research and Training Center (ISLRTC) released the first Indian Sign Language Dictionary, which has 3000 entries. In the second edition it is increased to 6000 terms on 27th Feb 2019 and 10000 terms in third edition on 17th Feb 2021. There is a lack of technical aid for Indian Sign Language which makes it difficult for signers to communicate with non-signers. New techniques and algorithms for quickly, precisely, and affordably recognizing sign language alphabets are now available due to advancements in machine learning and deep learning technologies.

Over the years, Deep Learning has been used in studies with the well-known sign languages. However, in most CNN model ReLU activation function is used as default till the fully connected layer and that has a major effect when it comes to read negative values. The major goal of this research is to identify the best activation function that suits for recognizing sign language with CNN model. The rest of the work is organized as follows. Section II describes a review on previous methods. Summary of dataset and methods used is given in section III. Section IV gives result with discussion and Conclusion is given in Section V.

II. Related Work

Shujun et al. [1], developed multitudinous spatiotemporal system for sign language identification. This paper proposed random sampling to select and adjust sequential images from video, which reduce the redundancy and increase the effectiveness of multimodal. D-shift Net is used to make primary motion features in a temporal stream and obtained over 96% on CSL dataset and above 60% on IsoGD. Wei Pan et al. [4], proposed ideal keyframe oriented clips sampling to detect important patterns from sign captured videos. Arif-Ul-Islam et al. [8] used orientation hash code and artificial neural network to recognize sign language in which proposed combination of ANN and hash code produce 95.8% accuracy.
Shalin Alom et al. [11], utilized a convolutional neural network to detect pattern from the sign number images and support vector machine to perceive digits in sign language and saw that the proposed model accomplished over 98% precision. Lance et al. [12], described a bidirectional sign language interpretation system using a convolutional neural network and achieved accuracy of 90%. Soma et al. [13] presented template matching technique to display equivalent text for the corresponding sign. Mengyi et al. [14], used a residual neural network for American Sign Language recognition which produced 99% accuracy.

YueSun et al. [18], used an extenics immune neural network to recognize important elements of Chinese Sign Language (CSL). Ilya Makarov et al. [19], compared several sign language three-syllable identification systems and proposed a deep convolutional neural network to recognize

Russian Sign Language (RSL) three-syllable hand sign images. Overall, accuracy is above 75%. Ilias papastratis et al. [20], utilized incessant sign language recognition involving Convolutional Neural Network for spatial element extraction, temporal convolution layers that has stacked 1D for short-term temporal modeling, and bidirectional long transient memory units for context learning.

Jie Huang et al. [25], used three-dimensional convolutional neural network for learning non-stationary features from unprocessed video to identify isolated Sign Language with an accuracy of 88% on their CSL dataset and 95% on ChaLearn14 dataset.

III. Materials and Methods

A. Dataset Describtion

One of the most crucial and important aspects of any investigation is data collection. It is crucial to gather information that is pertinent to the research and complies with its requirements. Readings from the data should duplicate every scenario, even the most extreme ones. These aids in making observations that are more complete and accurate. The data must be gathered with the utmost care because it serves as the basis for all study. In this experiment American Sign Language dataset from kaggle repository is used. The dataset contains American Sign for the Alphabets from A - Z except J and Z.

Fig. 1: American sign Alphabets

Though the size of the data is not evenly distributed among alphabets but it gives enough files with different angles for each character to carry out the research which is detailed in the below table.

Table 1: Number of samples for each label

Label	Size	Label	Size	Label	Size	Label	Size
A	539	G	345	N	293	T	301
B	541	H	364	O	374	U	286
C	387	I	360	P	221	V	337
D	379	K	319	Q	275	W	347
E	498	L	346	R	291	X	310
F	420	M	277	S	314	Y	318

There are 8442 RGB image files in the dataset out of which 80% is considered for training purpose and 20% for testing purpose.

B. Activation Function

The most crucial element in a neural network that determines whether or not a neuron will be engaged and moved to the next layer is the activation function which gives the ability for neural network to deal with non-linear problem. The output can be normalized using activation functions to fall between 0 and 1 or -1 and 1. Due to its differentiable property, it aids in the backpropagation process. Backpropagation updates the loss function, and the gradient descent curves are assisted in reaching their local minima by the activation function. The binary classification process can be readily carried out for single layer perceptron which can be seen in Figure 2.

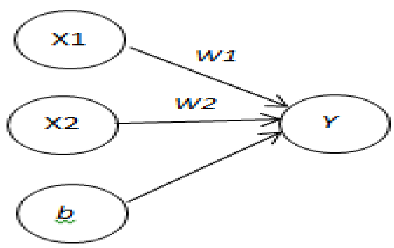


Fig. 2: Single Layer perceptron without Activation Function

From figure 2, *Y* can be defined as:

$$Y = w_1x_1+w_2x_2+b \quad (1)$$

When y=0, the line for linear classification can be obtained. Since the single layer perceptron cannot tackle the linear indivisibility problem, the multilayer perceptron can manage the multiclass problem based on the following equation.

$$y1 = \sum w_i x_i + b \quad (2)$$

However, a non-linear system classification problem cannot be solved by the classifier because it is fundamentally a linear equation, regardless of the combination. So that, activation function is added to the perceptron.

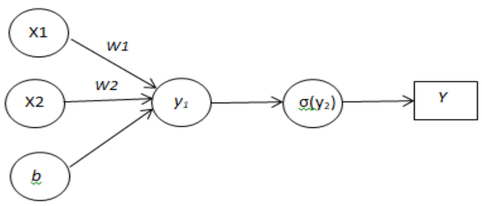


Fig. 3: Single Layer perceptron with Activation Function

The output of the above model can be defined as:

$$y_1 = w_1x_1 + w_2x_2 + b \quad (3)$$

$$Y = \sigma(y_1) \quad (4)$$

The neural network with activation function can handle non-linear classification problem by equation (3) and (4).

C. Common Activation Function Used in CNN

The key component of a deep neural network's architecture is its activation function, and common activation function includes: sigmoid, tanh, ReLu and Softmax but when it's come to Convolutional Neural Network most commonly used activation functions are variation of ReLu and Softmax. The ReLu function's limitation is that it can only address the issue of the gradient disappearing when the variable value is positive.

There are other activation functions such as LeakyReLu, PReLu, SELU, Swish, HardSwish, and Mish, which were also employed to address the issue of gradient disappearance when the variable value was negative. The curve variance of these can be seen in Figure 4. The following are the equation of these above activation functions respectively:

$$f(x) = \begin{cases} x & (x > 0) \\ a.x & (otherwise) \end{cases} \quad (5)$$

$$f(x) = \begin{cases} x & (x \geq 0) \\ a.x & (x \leq 0) \end{cases} \quad (6)$$

$$f(x) = \min(\max(0, x), 6) \quad (7)$$

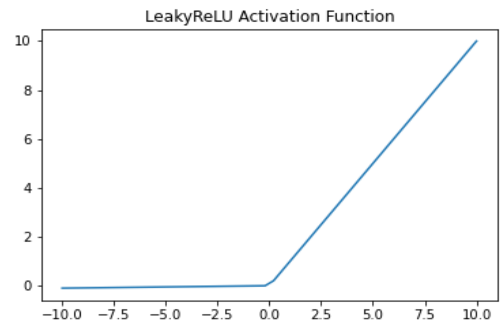
$$f(x) = \begin{cases} x & (x > 0) \\ \alpha(e^x - 1) & (x \leq 0) \end{cases} \quad (8)$$

$$f(x) = \mathrm{x} * \mathrm{sigmoid(x)} \quad (9)$$

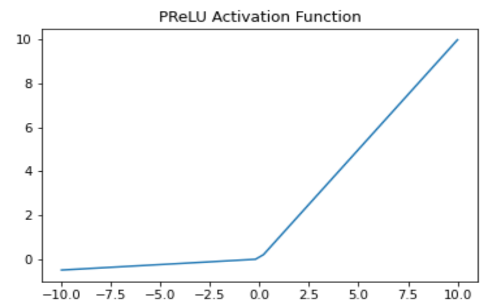
$$f(x) = \mathrm{x} * \mathrm{ReLU6(x + 3)/6} \quad (10)$$

$$f(x) = \mathrm{x} * \tanh(\mathrm{softplus(x)}) \quad (11)$$

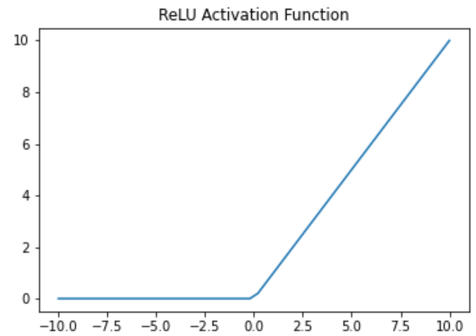
Some of the above listed variation functions outperformed the default ReLu activation function in terms of recognition outcomes on a sign language dataset. It is discovered through the comparison of various experimental findings that the Mish activation function's performance is steady and it increased the accuracy of the test set to a certain extent and also provides better convergence.



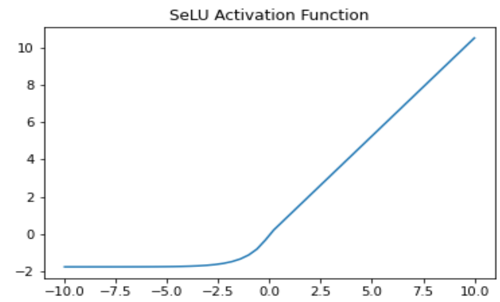
(a) The curve of LRelu function



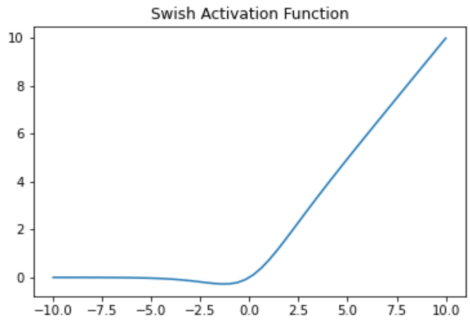
(b) The curve of PReLU function



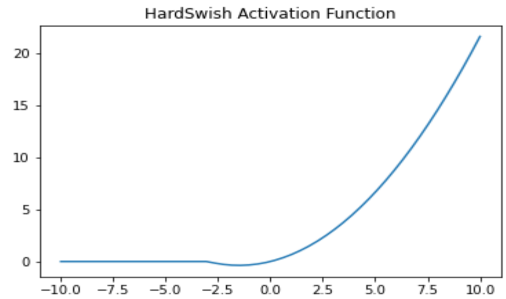
(c) The curve of Relu6 function



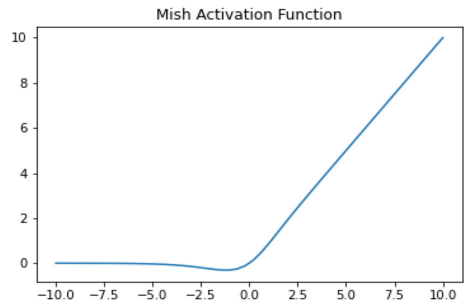
(d) The curve of SELU function



(e) The curve of Swish function



(f) The curve of HardSwish function



(g) The curve of Mish function

Fig. 4: The graph of above activation functions are LReLu, PReLu, SELU, Swish, HardSwish, and Mish.

D. Convolutional Neural Network Model

The proposed model is initialized with sequence of layers. Through the model, the information is transferred from the input layer to the hidden layer to the output layer. Convolution main goal is to use a feature detector to extract features from the input image and map those features onto a feature map, and maintain the spatial relationship between pixels.

Each CNN has two layers. The first layer in the architecture is a pair of convolutional layers with 32 filters and a 3x3 window size followed by a maxpool layer. Another set of convolutional layers with 64 filters, a max pooling layer and flatten layers are added at the end. After that a fully connected layer with 128 neurons ReLU activation function and an output layer with Softmax activation function are added. The starting convolutional layers takes an input image size of 64x64 where the final output layer has 24 neurons with respect to each class of ASL signs.

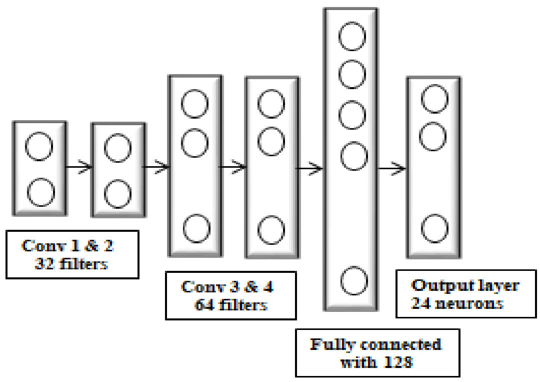


Fig. 5: The architecture diagram for CNN

IV.Results and Discussions

In order to create a highly accurate system that would be beneficial for real-time users, sign language recognition requires effective and robust data. The proposed system utilizes kaggle repository ASL dataset which contain sign images for 24 alphabet classes. After preprocessing, data is passed into CNN model for classification. Flowchart of proposed method is illustrated in Figure 6.

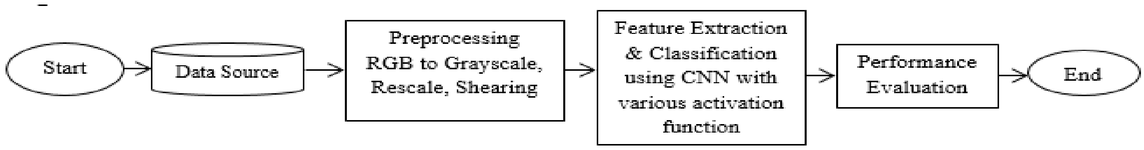


Fig. 6: Flow Chart of Proposed Model

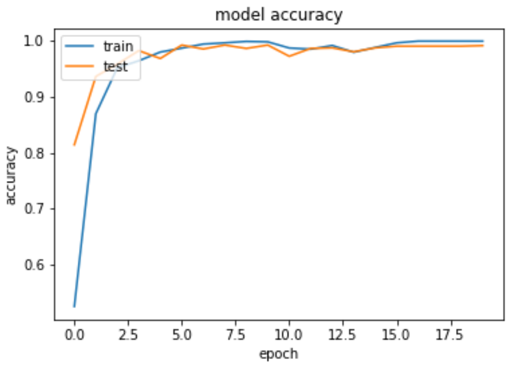


Fig. 7: Accuracy graph of CNN

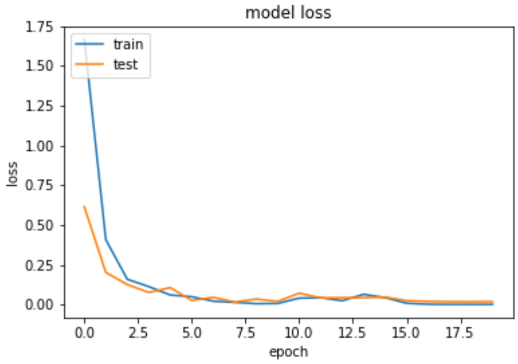


Fig. 8: Loss graph of CNN

Table 2: Accuracy of different activation function used in CNN for sign language recognition

Activation Function	ReLU	LReLU	PReLU	ELU	SELU	Swish	Hard Swish	Mish
Accuracy	94.85	96.97	94.75	92.43	93.98	95.25	98.72	99.63

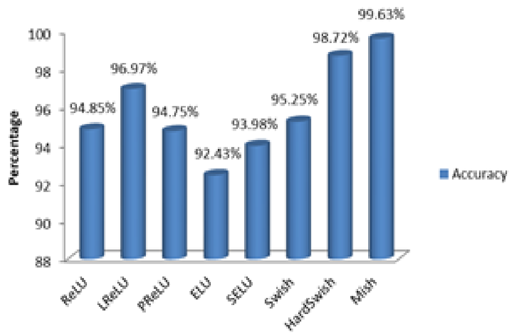


Fig.9: Performance of various activation function in CNN for sign language recognition

From the above table 2 it is clear that mish activation function outperform than other activation function with CNN for recognizing sing language. The accuracy rate produced by mish activation function is stable and which is 4.78%, 2.66%, 4.88%, 7.2%, 5.65%, 4.38% and, 0.91% higher than ReLu, LReLU, PReLU, ELU, SELU, Swish, HardSwish, and Mish activation function respectively. Figure 9 shows the confusion matrix for CNN with mish activation function with the learning rate 0.001

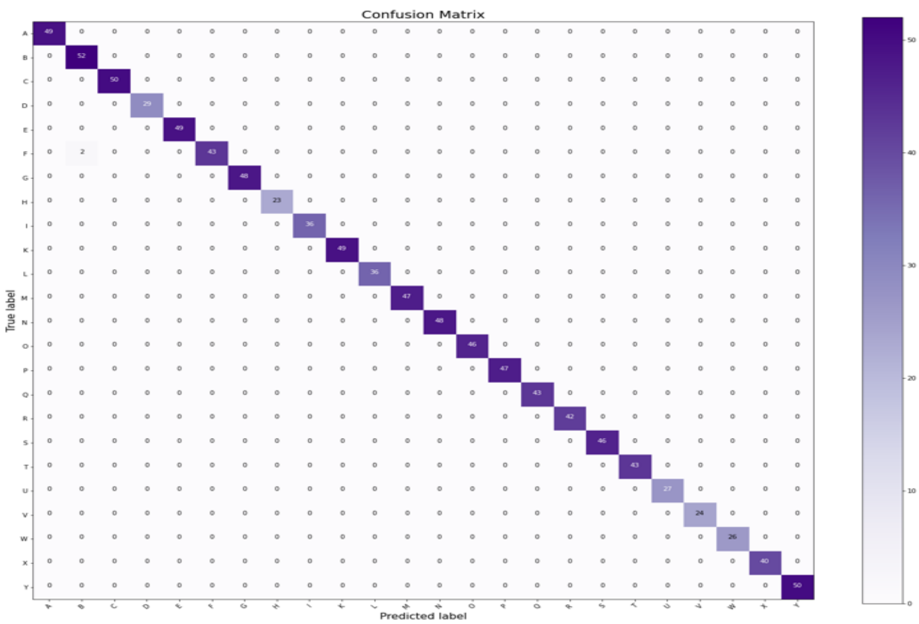


Fig. 10: Confusion matrix of CNN with Mish Activation Function

Table 3: Performance Assessment Table

Measure	Precision	Recall	F1 Score
CNN with Activation Function	99.06	99.02	99.28

Precision is defined as the proportion of predicted positive observations to all positive observations. The F1 score is the weighted average of precision and recall, whereas recall is the proportion of properly predicted positive labels to the total number of positive labels. The outcomes are displayed in Table 3.

V. Conclusion and Future Work

A common tool for image classification tasks is the CNN model. A crucial component of the convolutional neural network that can map out the non-linear characteristic is the activation function. This research investigates how the CNN model's activation function affects the recognition of sign language from the viewpoint of the activation function and it is found that CNN with mish activation function performs better for recognizing the sign language than the other activation function. In the future work, multi-model sign language will be recognized in real time using the proposed model.

References

1. Hemalatha S, Jayasumithra R, Priya Shalini R, A survey about deaf and dumb assisting device, International Journal of Creative Research Thoughts (IJCRT), vol. 8, pp. 1861-1864, 2020.
2. Pallavi A, Amrita I, Pranitha B.L, Darshini S, Paramesh R, Nirvatha Vadathi-An App to Assist Deaf and Dumb, International Journal of Engineering Research & Technology (IJERT), NCCDS-2020 Conference Procee-dings, Vol. 8, pp. 150-152, 2020.
   Kanwal Yousaf, Zahid Mehmood, Tanzila Saba, Amjad Rahman, Muhammad Rashid, Muhammad Altaf, Zhang Shuguang, A Novel Technique for Speech Recognition and Visualization Based Mobile Application to Support Two-Way Communication between Deaf-Mute and Normal Peoples, Hindwai Wireless Communication and Mobile Computing, https://doi.org/10.1155/2018/1013234,vol. 2018.
3. Shujun Zhang, Weijia Meng, Hui Li, Xuehong Cui, Mutlimodal Spatiotemporal Networks for Sign Language Recognition, IEEE Access, vol. 7, pp. 180270-180280, 2019.
4. Wei Pan, Xiongquan Zhang, Zhongfu Ye, Attention-Based Sign Language Recognition Network Utilizing Keyframe Sampling and Skeletal Features, IEEE Access, vol.8, pp. 215592-215602, 2020.
5. Anil Kumar D, Kishore P.V.V, Teja Kiran Kumar M, S3DRGF: Spatial 3-D Relational Geometric Features for 3-D Sign Language Representation and Recognition, IEEE Signal Processing Letters, vol. 26, pp. 169-173, 2019.
6. Tiago Oliveira, Nuno Escudeiro, Paula Escudeiro, Emanuel Rocha, Fernando Maciel Barbosa, The VirtualSign Channel for the communication Between Deaf and Hearing Users, IEEE Revista Iberomerciana de Technologias del Aprendizaje, vol. 14, pp.188 -195, 2019.
7. Arif-Ul-Isalm, Shamim Akhter, Orientation Hashcode and Artificial Neural Network Based Combined Approach to Recognize Sign Language. In: 21st international conference of Computer and Information Technology (ICCIT), doi:10:1109/ICCITECHN.2018.8631967, 2018.
8. Necati Cihan Camgoz, Oscar Koller, Simon Hadfield, Richard Bowden, Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation, In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), doi: 10.1109/CVPR42600.2020.01004, 2020, pp.10020-10030.
9. Amandeep Singh Dhanjal, Williamjeet Singh, Comparative Analysis of Sign Language Notation System for Indian Sign Language. In: Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), doi:10.1109/ICAPP.2019.8883009, 2019.
10. Shahin Alom, Jashid Hasan, Ferdous Washid, Digit Recognition in Sign Language Based on Convolutional Neural Network and Support Vector Machine. In: International Conference on Sustainable Technologies for Industry 4.0 (STI), doi:10.1109/STI47673.2019.9067999, 2019.
11. Lance Fernandes, Prathmesh Dalvi, Akash Junnarkar, Anisa Bansode, Convolutional Neural Network Based Bidirectional Sign Language Translation System. In: Proceedings of the 3rd International Conference on Smart System and Inventive Technology (ICSSIT), doi: 10.1109/ICSSIT 48917.2020.9214272, 2020.
12. Soma Shrenika, Myneni Madu Bala, Sign Language Recognition Using Template Match-ing Technique. In: International Conference on Computer Science,

Engineering and Applications (ICCSEA), doi: 10.1109/ICCESA49143.2020.9132899, 2020.
13. Mengyi Xie, Xin Ma, End-to-End Residual Neural Network with Data Augmentation for Sign Language Recognition. In: IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) , doi: 10.1109/IAEAC47372.2019.8998073, pp. 1629-1633, 2019.
14. Muth Mariappan H, Gomathi V, Real-Time Recognition of Indian Sign Language. In: Second International Conference on Computa-tional Intelligence in Data Science (ICCIDS), doi: 10.1109/ICCIDS.2019.8862125, 2019.
15. Yingying Wang, Yibin Li, Youn Song, Xuewen Rong, The influence of the Activation Function in Convolutional Neural Network Model of Facial Expression Recognition, Applied Science, https://doi.org/10.3390/ app10051897, 2020.
16. Kshitij Bantupalli, Ying Xie, American Sign Language Recognition using Deep Learning and Computer Vision. In: IEEE International Conference on Big Data (Big Data), doi: 10.1109/BigData.2018.8622141, pp. 4896-4899, 2018.
17. Yue Sun, Tintian Yuan, Junfen Chen, Rui Feng, Chinese Sign Language Key Action Recognition Based on Extenics Immune Neural Network. In: IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), doi: 10.1109/AEECA49918. 2020. 9213483, pp.187-191, 2020.
18. Ilya Makarov, Nikolay Veldyaykin, Maxim Chertkov, Aleksei Pokoev, Russian Sign Language Dactyl Recognition. In: 42nd International Conference on Telecommunica-tion and Signal Processing (TSP), doi:10.1109/ TSP.2019.8768868, pp. 726-729, 2019.
19. Ilias Papastratis, Kosmas Dimitropoulos, Dimitrios Konstantindis, Petros Daras, Continuous Sign Language Recognition through Cross-Modal Alignment of Video and Text Embeddings in a Joint-Latent Space, IEEE Access, vol. 8, pp. 91170-91180, 2020.
20. Sakshi Sharma, Sukhwinder Singh, Vision-based sign language recognition system: A Comprehensive Review, In: 5th International Conference on Inventive Computation Technologies (ICICT), doi: 10.1109/ICICT 48043.2020.9112409, pp. 140-144, 2020.
21. Andre J. Henry, Prangnat Chininthorn, A mobile survey for collecting data from Deaf people who use sign language for communication. In: IEEE Conference on Information Communication Technology and Society (ICTAS), doi: 10.1109/ICTAS50802. 2021.9394973, pp.79-84, 2021.
22. Tolba M.F, Elons A.S, Recent Developments in Sign Langauge Recognition Systems, In: 8th International Conference on Computer Engineering & Systems (ICCES), doi: 10.1109/ICCES.2013.6707157, pp.36-42, 2013.
23. Tianming Zaho, Jian Liu, Yan Wang, Hongbo Liu, Yingying Chen, Towards Low-Cost Sign Language Gesture Recognition Leveraging Wearables, IEEE Transactions on Mobile Computing, Vol. 20, pp.1685-1701, 2021.
24. Jie Huang, Wengang Zhou, Houqiang Li, Weiping Li, Attention based 3D-CNNs for Large-Vocabulary Sign Language Recogni-tion, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 29, pp. 2822-2832, 2019.
25. Vijayakumar, T. "Synthesis of Palm Print in Feature Fusion Techniques for Multimodal Biometric Recognition System Online Signature." Journal of Innovative Image Processing (JIIP) 3, no. 02 (2021): 131-143.
26. Smitha, T. V. A Study on Various Mesh Generation Techniques used for Engineering Applications, Journal of Innovative Image Processing 3, no. 2 (2021): 75-84.
27. Asutosh Hota, Risha Agarwal, Sai Sankar Gochhayat, Sandeep Mahapatra, Peformance Analysis of supervised Classification Models on Gesture Recognition of Single Handed Alphabet System of Indian Sign Language, International Journal of Advances in Electronics and Computer Science, Vol. 4, Issue.3, pp.92-97, Mar 2017.
28. Madhiarasan A, Partha Pratim Roy, A Comprehensive Review of Sign Language Recognition: Different Types, Modalities, and Datasets, Journal of Latex Class Files, Apr 2022.
29. Radha Shirbhate, Vedant Shinde, Sanam A, Pooja Boarkar, Mayuri Khandge, Sign Language Recognition Using Machine Learning Algorithm, International Research Journal of Engineering and Technology, Vol.7, Issue. 3, pp.2122-2125, Mar 2020.