Sustainable Hand Gesture Recognition for Speech Conversion, Empowering the Speech-Impaired

Sukanya Ledalla[1] *****, *G.Vijendar* Reddy[1], *Y Jeevan* Nagendra Kumar[1], *JOGINIPELLY* SHAILIKA[1] , *Minakshi* Rajput[2,]

[1]Department of Information technology, GRIET, Bachupally, Hyderabad, JNTUH, Telangana,India,500090.

[2]School of Applied and Life Sciences, Uttaranchal University, Dehradun, 248007, India

Abstract. A sustainable language disorder affects an individual's ability to reach out to others through speaking and listening. So utilizing sustainable hand gestures is among the most widespread means of non-verbal and visual communication used by people with speech disabilities worldwide. However, even though sustainable sign language is used everywhere by speech-impaired and hearing-impaired people, most of the populace who don't have any knowledge about sign language face difficulties in sustainably communicating with them. This sustainable problem requires better solutions that can successfully support communication for people with speech disabilities. This sustainable approach will reduce the communication gap for the speech-impaired population. There are many sustainable solutions in the market such as using sensors to make a sustainable device that gives a helpful output. But these sustainable solutions are expensive and not everyone can afford them. We are employing Convolutional Neural Networks to create a sustainable model that is trained on different gestures. This sustainable model enables speech-impaired individuals to convey their information using signs which get converted to human-understandable language, and sustainable voice is given as output. The sustainable hand gestures made are captured as a series of sustainable images which are processed using Python code. This sustainable endeavor introduces a solution that not only automates the identification of sustainable hand gestures but also transforms them into sustainable speech. By interpreting these recognized sustainable gestures, the corresponding recorded audio will be played sustainably. The focus of this sustainable paper is to offer accessibility, convenience, and safety to individuals with speech impairments in a sustainable manner. These sustainable individuals often experience societal discrimination solely due to their disabilities. This sustainable paper is aimed at innovating a sustainable device to help those without the knowledge of sign language sustainably communicate with the people who face difficulty in speech.

* Corresponding Author: ledalla.sukanya@gmail.com

1 Introduction

In the present sustainable era, the prevalence of automated technology has revolutionized various aspects of our daily tasks, significantly reducing the need for complex methodologies. However, it is disheartening to acknowledge that individuals with disabilities have not been able to fully reap the sustainable benefits of this automated environment. Particularly, the deaf and mute sustainable community continues to face significant challenges in their development and interactions with others due to their distinct mode of communication. Regrettably, the advancements in sustainable technology[11] have not adequately addressed the needs of individuals with disabilities, thus underscoring the sustainable importance of undertaking a paper that can bring about positive change for this marginalized group. Hence, the rationale behind the selection of the "Hand Gesture Recognition of Sign Language for Text and Voice Conversion" (HGRSLTV) sustainable program becomes apparent. This innovative sustainable initiative facilitates communication between deaf and mute individuals by enabling them to observe and trace the intricate sustainable movements of their hands. By leveraging the capabilities of a sustainable web camera, this sustainable research offers a practical solution in the form of hand motion detection and recognition.

2 Literature Survey

Akhilesh Pandey, Abhilasha Chauhan, Ashish Gupta, Vijay karnatak [1] stated that in our nation, approximately 2.78% of the population faces the challenge of being unable to speak. Communication is a fundamental aspect of human interaction, and throughout history, speech has been the primary means of communication. Despite advancements in science and technology that have improved the quality of life for many, there remains a segment of the population that struggles to find a solution to simplify their communication difficulties. Individuals who have difficulty speaking often rely on sign language, which is based on hand movements. This work introduces an innovative approach using a prototype of a Feed Forward Neural Network (FFNN) that can automatically recognize sign language, facilitating more effective communication between hearing, speech, or visually impaired individuals and those without such challenges. The system utilizes hand signal feature point extraction through a Feed Forward neural network. Additionally, a Hand Gesture Recognition with Voice Process system incorporating Hidden Markov Model (HMM) is employed to facilitate communication between individuals who cannot speak and those who can.
Amal Abdullah Mohammed Alteaimi and Mohamed Tahar Ben [2] Othman stated that the Hand Gesture Recognition (HGR) System provides an effective solution to facilitate communication between humans and computers[12] by eliminating the need for specialized input and output devices that can complicate interactions, particularly for individuals with disabilities. Hand gestures serve as a natural means of communication in human-computer interactions, and previous research has focused on using machine learning algorithms to accurately understand and recognize specific hand gestures. This research aims to create a strong hand gesture recognition model that achieves a perfect recognition rate of 100%. To realize this goal, a proposal is made for an ensemble classification model. This model combines multiple robust machine learning classifiers to enhance accuracy by incorporating diversity. The technique of majority voting is applied to amalgamate the accuracy scores generated by every classifier[13], leading to the conclusive classification result. Training involves a custom dataset containing 1600 images, representing ten distinct hand gestures. By combining the Canny edge detector and the histogram of oriented gradient method with the ensemble classifier, a notable recognition rate is accomplished. Experimental findings

validate the dependability of the suggested model, with Logistic Regression and Support Vector Machine reaching[14] 100% accuracy. Moreover, validation conducted using two publicly available datasets affirms the superior performance of the developed model compared to other comparative research endeavors.
Prof. Sonia Waghmare, Praveen Rathod, Ved Satdeve, Rohit Gaware, Ritesh Khapare [3] stated that Individuals are primarily dependent on visual and auditory cues to communicate back and forth. This interaction is a two-way interaction, which means that the speaker and the listener must both take part in the conversation. However, blind and deaf people are unable to communicate in this manner[15]. Their limitations make it difficult for them to communicate with others. Sign language is the primary method used by people who are deaf and blind. This form of communication is a set of gestures that are executed with the hands, arms, and face. The gestures are then conveyed to the other person. This form of communication is a visual and tactile form of communication that allows people who are deaf and blind to communicate with one another.
Tanaya Gadekar, Anjali Shette, Akash Gupta [4] stated that their objective is to develop a model capable of accurately recognizing and interpreting hand gestures and signs. The purpose of this paper is to facilitate communication between individuals who are naturally deaf or mentally disabled and others. We will train a model[6] specifically designed for sign language conversion, as well as a simpler gesture recognition model. Several implementation methods can be used, including K-Nearest Neighbors (KNN), Logistic Regression, Naïve Bayes Classification, Support Vector Machine (SVM) [7], and Convolutional Neural Network (CNN). After careful consideration, we have chosen to implement the CNN method due to its superior accuracy compared to other approaches.

The paper entails the development of a Python-based computer program that utilizes the CNN algorithm to train a model for hand gesture recognition. The program will compare input data with a pre-existing dataset created from the American Sign Language. Its output will convert sign language into text, enabling users to comprehend the signs conveyed by sign language speakers. The implementation of this model will be conducted within Jupyter Lab, an extension of the Anaconda platform, and appropriate documentation will be generated throughout the process.

To enhance the model's performance, we will also incorporate additional features such as converting inputs to black and white and utilizing background subtraction techniques when capturing images from the camera. By implementing a skin detection mask, the model will be able to function without requiring a plain background, thus enabling its usage with a basic camera and a computing device [16-20].

3 Methodology

In this paper the following steps are involved for hand gesture recognition to speech conversion

3.1. Dataset Preparation

Collect a labelled dataset of hand gesture images along with their corresponding speech labels. The dataset should include a diverse range of hand gestures that you want to recognize and convert into speech.

3.2. Image Pre-processing

Prepare the hand gesture images by improving their quality and eliminating any unwanted noise or extraneous details. Typical pre-processing actions encompass resizing, normalization, and filtration.

3.3. CNN Architecture

CNNs employ convolutional layers to detect features like edges and textures. Pooling layers downsample data to retain essential information. Deep layers capture complex patterns hierarchically. Fully connected layers translate features into final predictions. CNNs excel in image tasks, using parameter sharing and data augmentation.

3.4. Training

The CNN model undergoes training using the meticulously curated dataset. Throughout this training process, the model becomes adept at extracting pertinent features from the images depicting hand gestures and associating them with their respective speech labels. This proficiency is achieved by minimizing a specified loss function, such as categorical cross-entropy, utilizing methodologies like backpropagation.

3.5. Testing and Validation

Evaluate the effectiveness of the trained model by examining its performance on a separate test dataset. Employ metrics such as accuracy, precision, recall, and F1 score to gauge the efficiency of the hand gesture recognition system.

3.6. Speech Conversion

Once the CNN model identifies a hand gesture, establish a connection to the corresponding speech label. This can be achieved by associating each recognized gesture with a specific speech output using a lookup table or a mapping function.

A convolutional neural network (CNN) is a prominent form of artificial neural network extensively applied in deep learning, specifically tailored for processing visual information. These networks are often referred to as shift-invariant or space-invariant artificial neural networks (SIANNs) due to their distinctive architecture involving shared weights. They produce isovariant responses through the utilization of convolutional kernels or filters. CNNs are often associated with achieving invariance, but their downsampling operations can limit their ability to remain completely invariant. Convolutional neural networks (CNNs) have a broad range of applications across different fields, covering tasks like recognizing images and videos, suggesting systems, categorizing and segmenting images, interpreting medical images, processing natural language, interfacing with brain-computer technologies, and analyzing financial time-series information.

CNNs are derived from multi-layer perceptron’s but with modified connections. In contrast to fully connected networks where each neuron is connected to every neuron in the following layer, CNNs adopt a more localized and hierarchical structure. This helps prevent overfitting, which is a common problem in fully connected networks. CNNs employ regularization strategies like weight decay, skipped connections, and dropout, as they gradually construct more intricate representations by utilizing smaller and more straightforward patterns stored in their filters.

Taking cues from the structural arrangement observed in the visual cortex of living organisms, CNNs bear a strong resemblance to the receptive fields of individual cortical neurons. Within the visual domain, each neuron reacts to inputs within a confined region referred to as the receptive field. The partial overlap of receptive fields among numerous neurons results in a collective coverage of the entire visual field. CNNs utilize this principle, similar to other image classification algorithms, and require minimal pre-processing. Unlike traditional methods that rely on handcrafted filters, CNNs learn to optimize these filters (or features) through automatic learning. This ability to extract features without prior knowledge or human intervention is a significant advantage of CNNs.

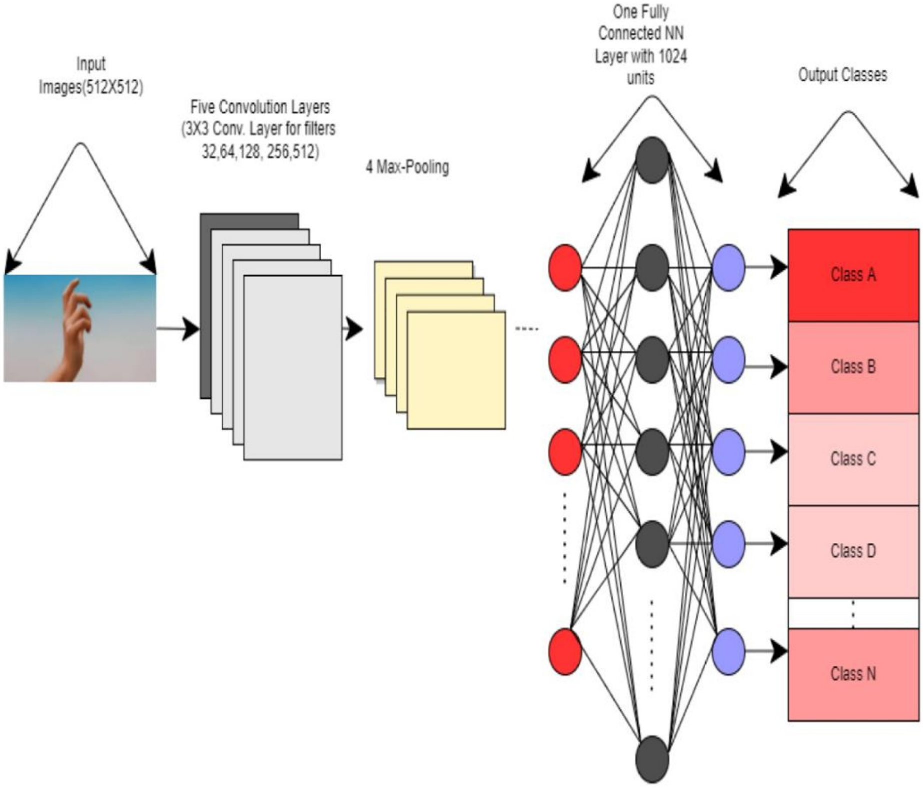
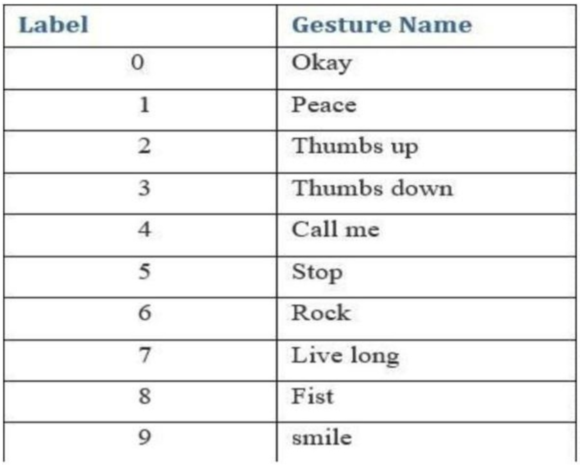


Figure 1. Work flow of the proposed System

4 Data

- The code we used does not explicitly have a dataset for generating output.

- Rather than using a collection of gesture images for training [9], we used a pretrained model which when provided with gestures identifies meaning of gesture.
- Whenever the gesture is shown ,'mpHands' module from Mediapipe is initialized to detect hand landmarks.
- The gesture recognition model is loaded using 'load_model' from 'tensorflow.keras.models.' The model is assumed to be saved in file named 'mp_hand_gesture'.
- The gesture recognition model which is trained with the help of some pretrained dataset is helpful for providing us the ouput.
- The class names for the gestures are loaded from a file named 'gesture.names'.
- In such a way, even though there is no direct dataset [8] that is being used in our code a pretrained dataset is used.



Label	Gesture Name
0	Okay
1	Peace
2	Thumbs up
3	Thumbs down
4	Call me
5	Stop
6	Rock
7	Live long
8	Fist
9	smile

Figure 2. Hand Gesture Labeling

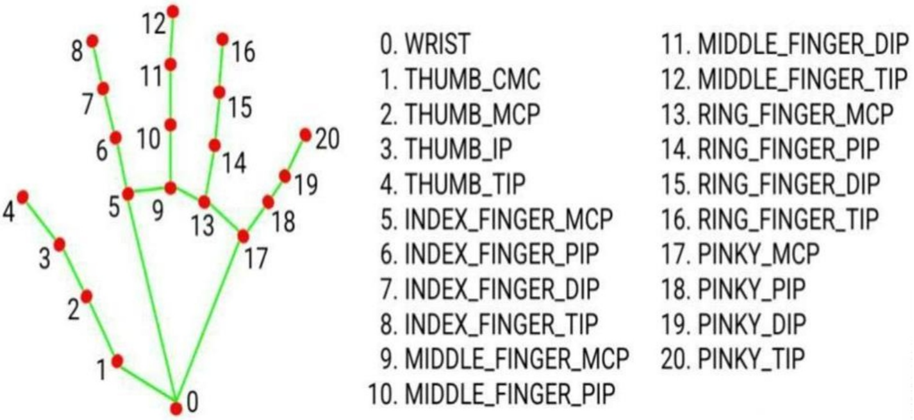


Figure 3. Hand Gesture Points

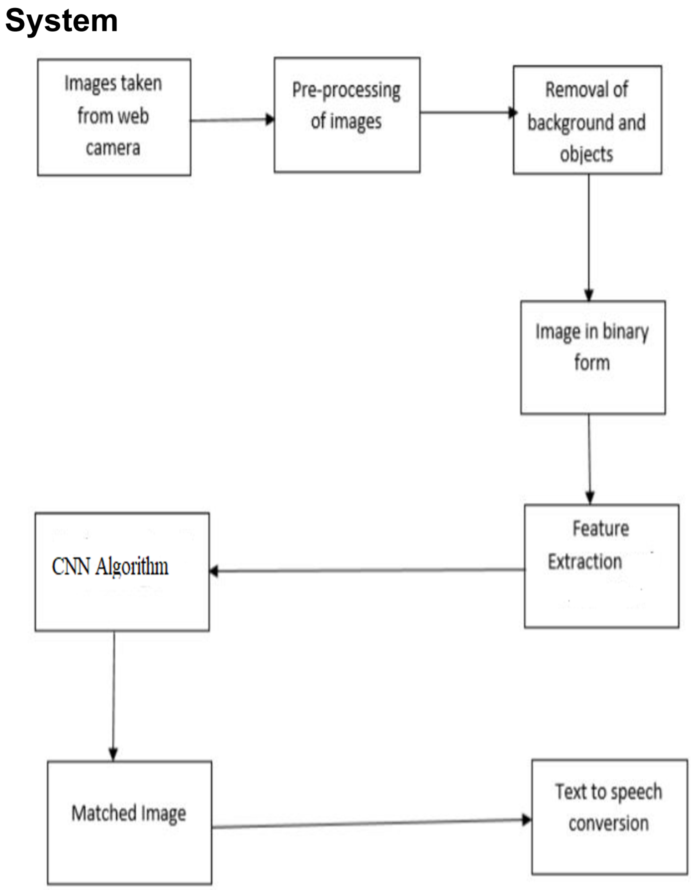


Figure 4. System Architecture

5 RESULT ANALYSIS

Utilizing MediaPipe for the machine learning-based recognition of hand gestures resulted in impressive outcomes, as depicted in the subsequent illustrations.

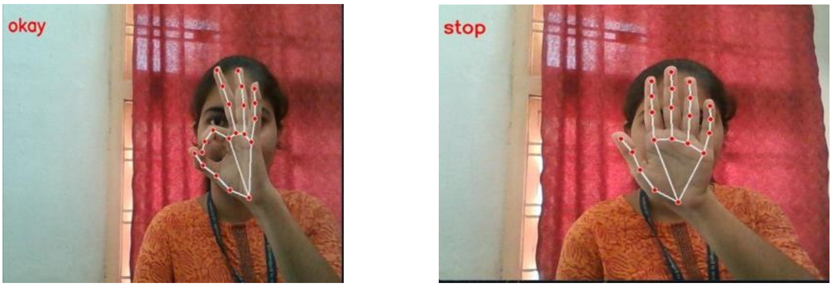
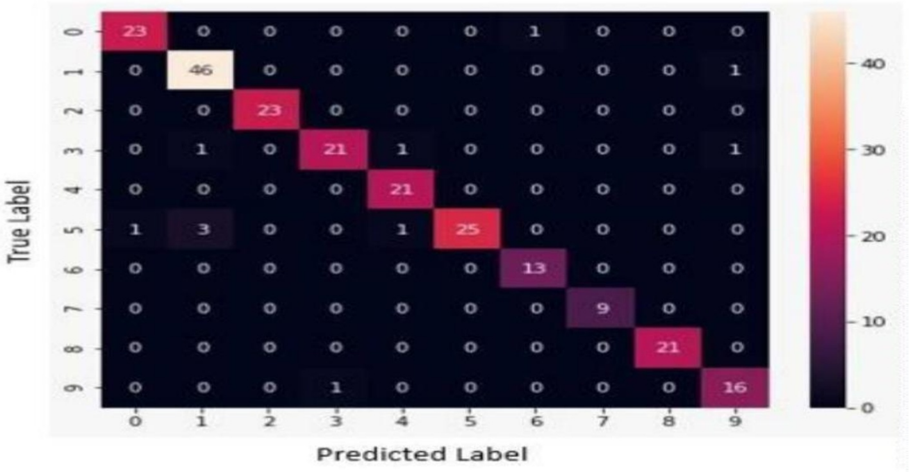
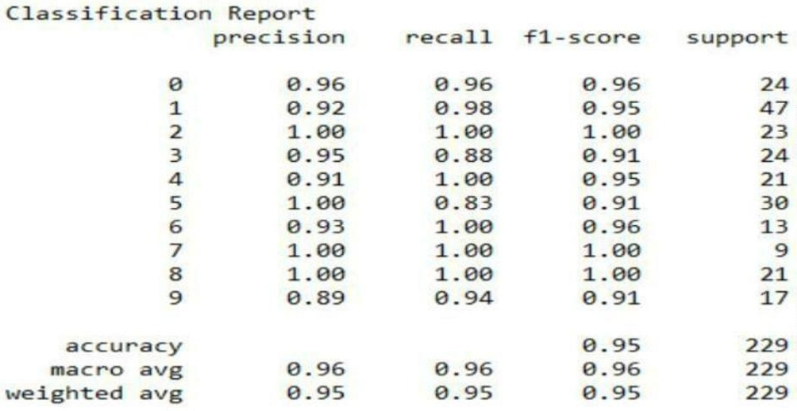


Figure 5. Sample Hand Gesture



	0	1	2	3	4	5	6	7	8	9
0	23	0	0	0	0	0	1	0	0	0
1	0	46	0	0	0	0	0	0	0	1
2	0	0	23	0	0	0	0	0	0	0
3	0	1	0	21	1	0	0	0	0	1
4	0	0	0	0	21	0	0	0	0	0
5	1	3	0	0	1	25	0	0	0	0
6	0	0	0	0	0	0	13	0	0	0
7	0	0	0	0	0	0	0	9	0	0
8	0	0	0	0	0	0	0	0	21	0
9	0	0	0	1	0	0	0	0	0	16

Figure 6.Confusion Matrix



```
Classification Report
              precision    recall  f1-score   support

           0       0.96      0.96      0.96        24
           1       0.92      0.98      0.95        47
           2       1.00      1.00      1.00        23
           3       0.95      0.88      0.91        24
           4       0.91      1.00      0.95        21
           5       1.00      0.83      0.91        30
           6       0.93      1.00      0.96        13
           7       1.00      1.00      1.00         9
           8       1.00      1.00      1.00        21
           9       0.89      0.94      0.91        17

    accuracy                           0.95       229
   macro avg       0.96      0.96      0.96       229
weighted avg       0.95      0.95      0.95       229
```

Figure 7.Classification Report

In the realm of machine learning, the evaluation of a model's efficiency frequently involves the utilization of a confusion matrix. In Python, this matrix can be produced using OpenCV. To appraise the proficiency of our model in generating meaning from hand gestures, we collected experimental datasets and produced a confusion matrix to ascertain its level of accuracy.

6 Conclusion and Future scope

This sustainable project focused on converting hand gestures into speech for individuals with speech impairments, we employed Python and OpenCV to construct a sustainable hand gesture recognition system. MediaPipe and Tensorflow frameworks were used for the detection and sustainable gesture recognition aspects, respectively. Through this sustainable endeavor, we gained knowledge about popular machine learning algorithms and essential libraries of Python which would help to build useful and sustainable projects. The sustainable project serves as a clear demonstration of how CNNs can effectively address computer vision challenges with high accuracy in a sustainable manner. We successfully developed a fingerspelling sign language translator achieving a perfect and sustainable accuracy rate. With the creation of essential sustainable datasets and CNN training, the sustainable project has the potential to expand its scope to encompass additional sustainable sign languages.

References

1. L. Ku,W. Su, P. Yu and S. Wei, “A real-time portable sign language translation system,” 2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS), Fort Collins, CO, 2015, pp. 1-4, doi: 10.1109/MWSCAS.2015.7282137.
2. S. Shahriar et al., “Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning,” TENCON 2018 2018 IEEE Region 10 Conference, Jeju, Korea (South), 2018, pp. 1168-1171, doi: 10.1109/TENCON.2018.8650524.
3. M. S. Nair, A. P. Nimitha and S. M. Idicula, “Conversion of Malayalam text to Indian sign language using synthetic animation,” 2016 International Conference on Next Generation Intelligent Systems (ICNGIS), Kottayam, 2016, pp. 1-4, doi: 10.1109/ICNGIS.2016.7854002.
4. M. Mahesh, A. Jayaprakash and M. Geetha, “Sign language translator for mobile platforms,” 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, 2017, pp. 1176-1181, doi: 10.1109/ICACCI.2017.8126001.
5. S. S Kumar, T. Wangyal, V. Saboo and R. Srinath, "Time Series Neural Networks for Real Time Sign Language Translation,” 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, 2018, pp. 243-248, doi: 10.1109/ICMLA.2018.00043.
6. Khalil Bousbai, Mostefa Merah. “A Comparative Study of Hand Gestures Recognition Based on MobileNetV2 and ConvNet Models”’ 2019 6th International Conference on Image and Signal Processing and their Applications (ISPA), 2019
7. Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In AAAI (pp. 4278- 4284)
8. Madhu Bala Myneni, L V Narasimha Prasad, J Sirisha Devi (2017). In A Framework for Sementic Level Social Sentiment Analysis Model. Journal of Theoretical and Applied Information Technology
9. Medel, J. R., & Savakis, A. (2016). Anomaly detection in video using predictive convolutional long short-term memory networks. arXiv preprint arXiv:1612.00390
10. J Sirisha Devi, Siva Prasad Nandyala, P Vijaya Bhaskar Reddy (2019). A Novel Approach for Sentiment Analysis of Public Posts. In Innovations in Computer Science and Engineering
11. Raju, NV Ganapathi, and Someswara Rao Chinta. "Region based instance document (rid) approach using compression features for authorship attribution." Annals of Data Science 5.3 (2018): 437-451.

12. Raju, NV Ganapathi, V. Vijay Kumar, and O. Srinivasa Rao. "AUTHORSHIP ATTRIBUTION OF TELUGU TEXTS BASED ON SYNTACTIC FEATURES AND MACHINE LEARNING TECHNIQUES." Journal of Theoretical & Applied Information Technology 85.1 (2016).
13. Prediction of diabetes using machine learning Jeevan Nagendra Kumar, Y. Kameswari Shalini, N. Abhilash, P.K. Sandeep, K. Indira, D. International Journal of Innovative Technology and Exploring Engineering, 2019, 8(7), pp. 2547–2551
14. Brain Tumors Classification System Using Convolutional Recurrent Neural Network V. Akila, P.K. Abhilash, P Bala Venakata Satya Phanindra, J Pavan Kumar, A. Kavitha E3S Web Conf. 309 01075 (2021) DOI: 10.1051/e3sconf/202130901075
15. H kanaka Durga Bella, DR.S. Vasundra ,CSE, JNTUA,"A study of security threats and attacks in Cloud Computing " IEEE-4th International conference on smart systems and inventive technology (ICSSIT-2022), DOI: 10.1109/ICSSIT53264.2022.
16. Prasanna Lakshmi, K., Reddy, C.R.K. A survey on different trends in Data Streams (2010) ICNIT 2010 - 2010 International Conference on Networking and Information Technology, art. no. 5508473, pp. 451-455.
17. Jeevan Nagendra Kumar, Y., Spandana, V., Vaishnavi, V.S., Neha, K., Devi, V.G.R.R. Supervised machine learning Approach for crop yield prediction in agriculture sector (2020) Proceedings of the 5th International Conference on Communication and Electronics Systems, ICCES 2020, art. no. 09137868, pp. 736-741.
18. Sankara Babu, B., Suneetha, A., Charles Babu, G., Jeevan Nagendra Kumar, Y., Karuna, G. Medical disease prediction using grey wolf optimization and auto encoder based recurrent neural network (2018) Periodicals of Engineering and Natural Sciences, 6 (1), pp. 229-240.
19. Nagaraja, A., Boregowda, U., Khatatneh, K., Vangipuram, R., Nuvvusetty, R., Sravan Kiran, V. Similarity Based Feature Transformation for Network Anomaly Detection (2020) IEEE Access, 8, art. no. 9006824, pp. 39184-39196
20. smart IoT based system for monitoring and controlling the sub-station equipment," Internet of things, vol. 7, p. 100085, 201smart IoT based system for monitoring and controlling the sub-station equipment," Internet of things, vol. 7, p. 100085, 2019