ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Bird Species Identification using Audio Processing and AlexNet Neural Network

Sanjay Gandhi Gundabatini[1], Professor, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
Sangam Sai Sri Vinay Reddy[2], Thota Chandrika[3], Somarouthu Kaarthikeya Pavana Kumaar[4], Shaik Siddik[5], Torlikonda Satya Akhil[6]
[2,3,4,5,6] UG Students, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
[1] sanjaygandhi.g@vvit.net,[2]sssvinayr@gmail.com,
[3]chandrikathota1005@gmail.com,[4]kaarthikeyapavanakumaar@gmail.com,[5]shaiksiddik630@gmail.com, [6]satyaakhil1602@gmail.com

DOI: 10.48047/IJFANS/V11/I12/173

Abstract

This research involved identifying the birds with the help of audio recorded from the real world environment. Most of the methods use images for detection of birds. But, some species may be similar to see. Hence, we took audio as the basis for classification. The audio frequency was plotted as spectrogram and it was inspected to extract the patterns and classify the bird. Legacy practices involved manual inspection of spectrogram that are plotted on the frequency of audio signals. But, this is time taking process and often produces inaccurate results. Hence, we created a computerized process to inspect the spectrogram. The computer learns from patterns of spectrogram during the training process and learns to detect a new and unseen audio. This entire procedure involved two crucial phases. The first stage of the process was to create a dataset with audio files collected from websites like Xeno-canto org that includes all sound recordings of birds. In this research work, we considered 4 species of wood pecker in the Germany region. Hence, we have collected approximately 120 recordings for each species, thus a total of 500 recordings were collected. The collected sounds undergone a series of pre-processing phases like reconstruction, framing, and silence removal, pre-emphasis for removing any noises like human actions, wind sounds, tree sounds. For every processed sound clip, the spectrogram is plotted and it was given as input to a neural network that is in the second stage, which in turn detects the recording at the end. Since the image was given as input, we used Convolutional Neural Network (CNN) which is a best neural net in deep learning for text and image based tasks. The CNN categorizes sound clip and determines the species of bird based on input features. A model was created and put into practice.

Keywords: Convolution, Neural Network, Spectrogram, Audio processing, Noise removal, Deep Learning, Classification.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Introduction

IUCN estimates that there are close to 10,000 bird varieties of that can be found in a vast range of environments, from Brazilian jungle to the cold beaches of Antarctica [1]. These species are vital to an ecosystem's well operation and exhibit incredible diversity in terms of behaviour and form. But recent human activities that range from encroaching on their ecosystems to completely destroying those habitats have jeopardised this wonderful biological diversity, and when combined with natural calamities and climate change, many bird species have been undergone extinction. An estimated 13% of all bird species, or close to 1,370 species, are thought to be in danger of going extinct. Despite the fact that many different bird species are frequently observed, people have a hard time identifying them. The objective of this work is to create computerised methods for classifying different species of birds by relying on their noises.

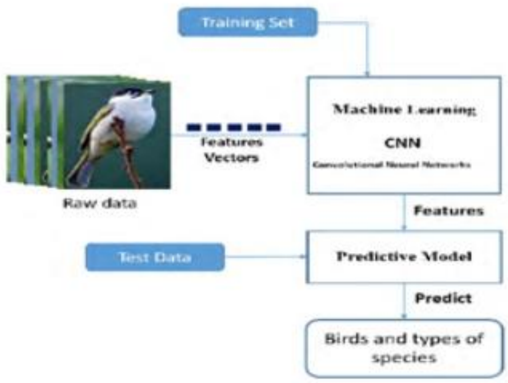

Fig.1: process flow figure

The rapid changes in bird population has been a major concern in recent times by many environmental organizations. However, acquiring information about different types of birds is a labor-intensive and expensive process that demands a lot of human effort. In this scenario, we need a good tool for preparing data is required. In this aspect, determining which bird species a particular image belong to, plays a vital role in determining which categories it belongs to. [14-22]

Bird species identification is determining which category a particular species of bird falls under using a photograph. Images, audio, and videos can all be used to identify different bird species. It is possible to identify birds by listening for their distinctive sounds thanks to an audio processing technique. However, handling of such data becomes increasingly complicated because of the mixed sounds in the condition, such as creepy crawlies, real-world items, and so forth. Images typically aid in information discovery more than sounds or recordings. In this perspective, utilizing an image rather than voice or video to categorize birds is preferred. Bird species identification is a difficult issue for both humans and

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

computational processes that perform such a task automatically. But, for more accuracy, despite of the above stated complications, we proceeded to audio based classification.

Literature Survey

A. Identification of Bird Species from Their Singing:

One of the enigmatic phenomena that we are learning more about as time goes on is biodiversity. The BIOTOPE organization has launched an initiative to aid in the process of comprehending nature. The group's open source platform will help to gather information by identifying different types of birds from their humming. The fundamental concept of this work is to recognize different types of birds by their songs or calls. The research discussed in this article investigates an autonomous system for classifying various bird species based on their songs and calls. This will make it easier to catalogue and identify new species or subspecies. Additionally, it will support research on bird behaviour and ecology and help with population surveillance. In this work, we put into practise a system that can find different bird species from audio recordings. According on the ROC performance analysis of the proposed system, Random Decision Tree may produce 0.96 area under the curve.

B. FFT Based Automatic Species Identification Improvement with 4- layer Neural Network:

An automated approach for identifying species has been devised in this paper. Recorded data was automatically split, analyzed, had features removed, and recognized. A 4-layer neural network-based feature quantity technique based on FFT and frequency band power derivative is provided. Wild bird species identification using a 4-layer neural network based on reliable data has been tested, and the results have shown promise.

C. Feature set comparison for automatic bird species identification:

In order to identify the species of a bird from itsrecordedaudio song, an issue known as automated bird species identification must be solved. Given that it is an indirect, non-invasive method of evaluation, this is an ingenious way to keep track of ecosystem biodiversity. This study compares various feature sets that describe the audio characteristics of the audio stream in various ways with machine learning methods like neural networks and SVM. Three different bird species' recorded songs are used as the subject of experiments. The experimental results demonstrate that it is possible to achieve highly encouraging results in automated bird species identification. The results compare the performance of the feature sets and several classifiers.

D. Detecting bird sounds in a complex acoustic environment and application to bio-acoustic monitoring:

Bird population trends are a crucial indication for environmental protection, but determining these trends requires a lot of time and human effort. Automated approaches for the detection of bird vocalizations are now conceivable through enormous advancements

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

in audio signal processing and pattern recognition. These tactics can be used to support population censuses. A study that examined the viability of automatic bird song detection in conjunction with bird monitoring was highlighted in this task. Describing new methods for the automatic habitat mapping of the vocalizations of two bird species that are in danger of extinction was given high importance. These techniques work by identifying temporal trends in a certain frequency band that are specific characteristics of the species. The noise that is present in real-world audio scenarios is specifically suppressed. Insights demonstrate that high recognition rates with a manageable number of false positive detection are still feasible under real-world recording situations.

E. Automated sound recording and analysis techniques for bird surveys and conservation:

Automated sound recording and analysis of bird sounds need to be used more and developed further. Since birds are essential to the health of ecosystems, science and conservation efforts would substantially benefit from methods that make avian monitoring more effective and precise. We give a briefing on popular methods used to identify and categorize bird sounds automatically, as well as a summarized view of the hardware approaches to automated sound recording. An interface for a scheduling timer to run a standalone recorder, a recording device that is programmable, and a computer with single board are three examples of the three basic kinds of hardware solutions we compare and contrast for automating sound recording. We also give examples of the two basic methods using wave guides and tiny arrays of microphone elements for enhancing microphone performance for automated recorders.

F. Deep Learning Based Audio Classifier for Bird Species

Problem Identification

Automatically identifying bird cries from continuous recordings gathered from the environment would be a huge advancement in the study methodologies. As noise or clipping are frequently present in these recordings, reliable automated processes must be used instead of manual, conventional techniques. Automated systems are required because manual spectrogram examinations are frequently mistaken and typically require a large number of professionals to employ specialized techniques.

Disadvantages:

1. Since manual checks of the spectrograms are frequently prone to error and typically involve many experts and esoteric methodologies, automated systems are necessary.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Methodology

In this paper, a method for automating the entire process of identifying bird sounds using sound processing and convolutional neural networks is proposed. The initial step entails compiling all of the sound recordings into a database. Then, sound pre-processing methods like pre-emphasis, framing, silence removal, and reconstruction are applied to these recordings.

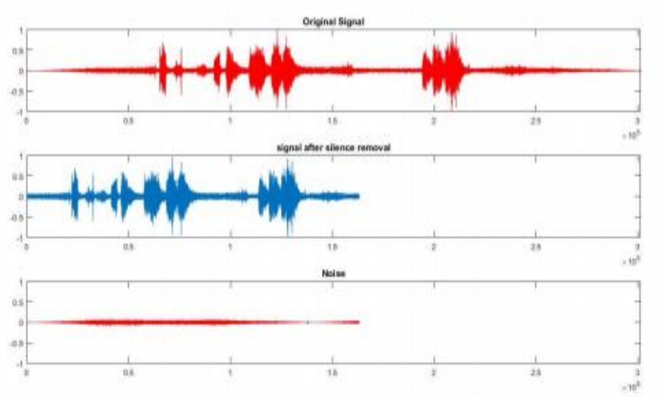

Fig 2: silence removal

Pre-emphasis

It is a technique to process signals that involves improvising the frequency components of an audio signal. It amplifies high audio components and diminishes the intensity of low frequency components.

Framing

This process divides audio signal into small and manageable parts. It assists in processing the signals with ease. But, very short duration of frames results in higher computational burden.

Silence removal

It is an approach to detect and remove the periods of silence removal from an audio signal. One common approach followed is to set a threshold frequency. If any signal is detected below this frequency, it is removed from the sample. This threshold is set based on the averages of all frequencies in the signal.

Reconstruction

This process involves digitally encoded audio signal back into its original wave form. It generates continuous wave form from the digital signal which matches the nearest frequency.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

For each sound clip, spectrograms are produced, and these spectrograms are fed into a CNN that has been trained on a GPU. For the trained CNN model, a real-time implementation model was developed.

Advantages:

1. This system has a 98% accuracy rate when classifying bird species based on the spectrogram image created from their sounds.
2. This precision was attained by taking into account both human voices and avian sounds.
3. By fine-tuning the performance settings, the precision can be increased even more.

Process flow

- Data exploration: this module loads data into the system.
- Processing: this module will read and process the data.
- Train_test data split: to divide the dataset into train and test data
- Generating the Model: Building the model Alexnet with CNN.

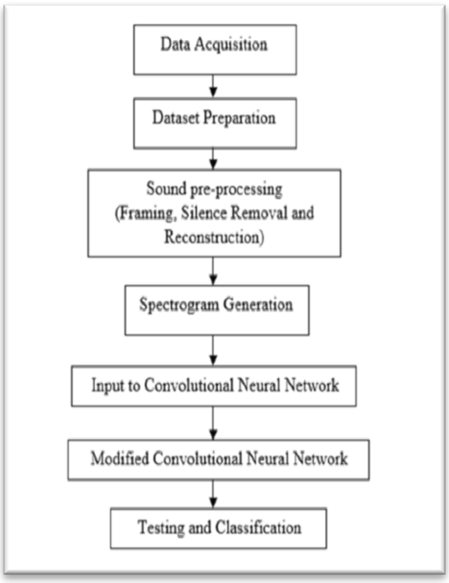

Fig.3: System architecture

- Predict accuracy: Using this module we will get accuracy, precision, recall and FSCORE
- User input: To take unseen data as input to test the model.

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Implementation

To identify or predict the species of bird, we are using the AlexNet CNN model, and in order to do so, we are using two different modules. The first module is used to read bird audio files, after which audio framing and noise removal are applied, and finally spectrogram features are extracted from the processed audio file. Applying the pre-trained CNN model from AlexNet on the extracted features in the second module to create a bird species recognition model. using the datasets of XENO CANTO bird audio files, which incorporates noise signals, to train the model.

ALEXNET:

A special type of neural network called a convolutional neural network (CNN) is used for supervised learning. CNNs are useful for processing images, natural language, and other cognitive tasks. AlexNet is the first neuralnet to make use of GPU training. 5 convolutional layers (3 are max-pooling layers, 2 are normalization layers), and we added 3 extra layers with (9,9), (7,7) and (6,6) number of neurons respectively, 2 layers that are fully connected, and a Softmax layer forms required architecture.

A multi-layer perceptron with a specific architecture for identifying two-dimensional picture data is the convolution neural network algorithm. An input layer, a convolution layer, a sample layer, and an output layer are among its four layers.

The sample layer and convolution layer in a deep network architecture could have more than one. Convolution neural network techniques do not have the same limitations as the Boltzmann machine in that each neuron just needs to sense the local portion of the image rather than the entire image, as is the case with the Boltzmann machine.

Additionally, each neuron's parameters are specified to be the same, including the sharing of weights and the use of identical convolution kernels for the deconvolution picture for each neuron.

The central elements of CNN are the local receptive field, weight sharing, and time- or space-based sub sampling with the aim of feature extraction and training parameter reduction.

The benefit of the CNN algorithm is that it learns intuitively from training data rather than explicitly extracting features. The network can learn in parallel and become less complex because the same neuron weights are on the surface of the feature mapping. utilizing a sub-sampling framework based on deformation displacement, scale, and time resilience.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Network topology and input data may go well together. It offers special benefits for picture processing. These actions are part of the Convolution Neural Network:

Convolution Layer:

It is the basic building block of CNN. It performs mathematical operation known as convolution which handles multi dimensional data like images, sound waves. This operation is useful to extract features from the given data. This layer consists of filters which are responsible for identifying patterns from the data. No. of filters in the layer is a hyper parameter which can be tuned for better control,performance.

In addition to convolution operation, it includes other operations like batch normalization and activation functions which will aid in improvising the stability of the network.

Pooling Layer:

Pooling layer is responsible for compressing the input data while holding the necessary features in it. This helps in reducing computational efforts and over-fitting of the network. Various types of pooling are:

- Max Pooling
- Mean Pooling
- Sum Pooling

Fully Connected Layer:

It is a layer in CNN whose individual node is connected to every other node in another layer. They are last layers in the network. . No. of neurons in the layer is a hyper parameter which can be tuned for better control, performance.

Results

```
In [3]: def getID(name): #function to get ID of the bird as label from species names
            index = 0
            for i in range(len(labels)):
                if labels[i] == name:
                    index = i
                    break
            return index
        labels = []
        for root, dirs, directory in os.walk("NoiseCleanAudio"):#now loop all files and get labels and then display all birds names
            for j in range(len(directory)):
                name = os.path.basename(root)
                if name not in labels:
                    labels.append(name)
        print("Bird Species found in dataset")
        print()
        for i in range(len(labels)):
            print(labels[i])

        Bird Species found in dataset

        Buff-throated Woodcreeper
        Lesser Woodcreeper
        Olivaceous Woodcreeper
        Yellow-olive Flatbill
```

Fig.4: Dataset

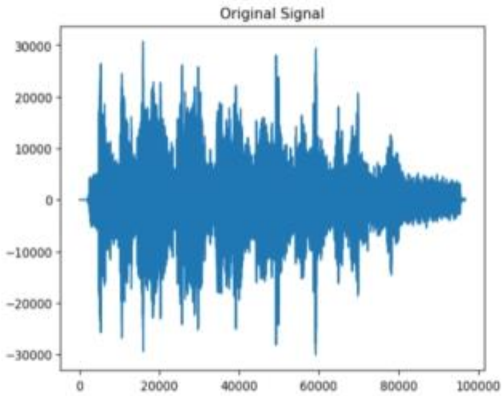

Fig.5: Audio file with noise

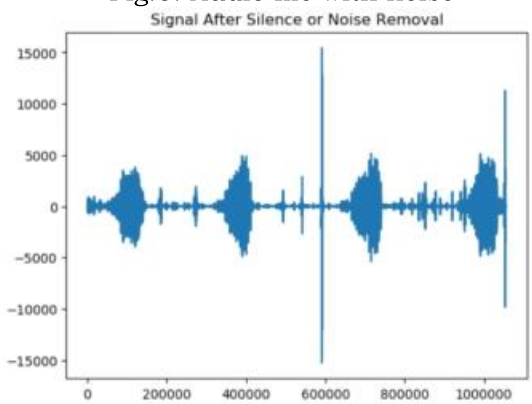

Fig.6: Audio without noise

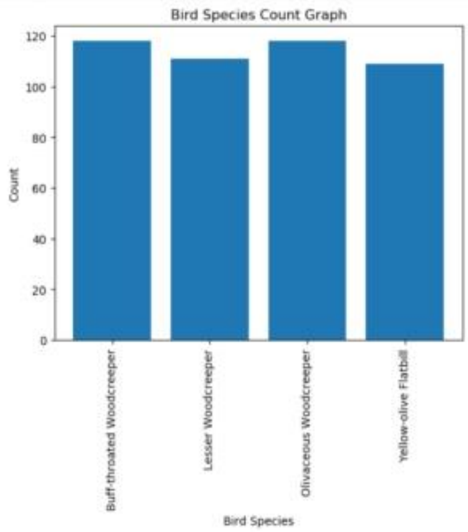

Fig.7: graph of various birds species and their count in the dataset

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

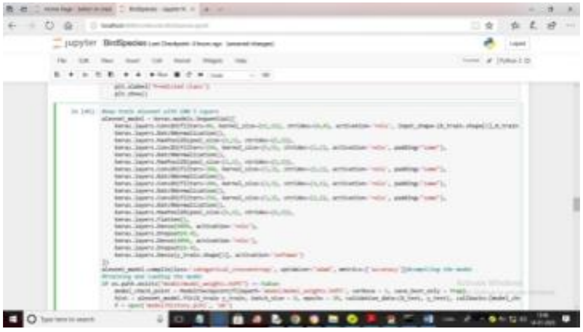

Fig.8: Model trained

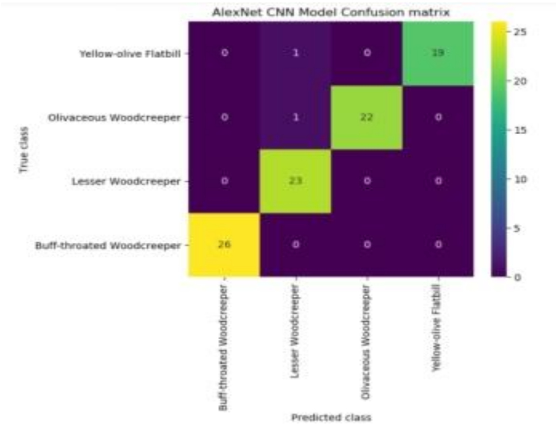

Fig.9: Confusion matrix

Fig.10: Prediction for test data

Conclusion

The system has a 98% accuracy rate when classifying bird species based on the spectrogram image created from their sounds. This precision was attained by taking into account both human voices and avian sounds. By fine-tuning the performance settings, the precision can be increased even more.

Limitations and Future Scope

- In the future, a mobile app could be created and released that would let consumers utilize their mobiles as handheld devices for bird sound analysis and prediction.
- The model would be trained on larger datasets to support wide range of species using big data techniques.
- We could use the advanced models that will be released in future which can efficiently handle large datasets and make the training process easier.

References

[1] Sujoy Debnath, Partha Protim Roy , Amin Ahsan Ali, M Ashraful Amin" Identification of Bird Species from Their Singing", 5th International Conference on Informatics, Electronics and Vision (ICIEV), 2016

[2] Rong Sun, Yihenew Wondie Marye, Hua-An-Zhao "FFT Based Automatic Species Identification Improvement with 4- layer Neural Network", 2013 13th International Symposium on Communications and Information Technologies (ISCIT)

[3] F. Briggs, B. Lakshminarayanan, L. Neal, X. Z. Fern, R. Raich, S. J. Hadley, A. S. Hadley, and M. G. Betts, "Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach," The Journal of the Acoustical Society of America, vol. 131, no. 6, pp. 4640–4650, 2012.

[4] M. T. Lopes, A. Lameiras Koerich, C. Nascimento Silla, and C. Alves Kaestner, "Feature set comparison for automatic bird species identification," in Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on. IEEE, 2011, pp. 965–970.

[5] R. Bardeli, D. Wolff, F. Kurth, M. Koch, K.-H. Tauchert, and K.-H. Frommolt, "Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring," Pattern Recognition Letters, vol. 31, no. 12, pp. 1524–1534, 2010.

[6] T. Scott Brandes, "Automated sound recording and analysis techniques for bird surveys and conservation," Bird Conservation International, vol. 18, no. S1, pp. S163–S173, 2008.

[7] A. Harma, "Automatic identification of bird species based on sinusoidal modeling of syllables," in Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on, vol. 5. IEEE, 2003, pp. V–545.

[8] A. Selin, J. Turunen, and J. T. Tanttu, "Wavelets in recognition of bird sounds," EURASIP Journal on Applied Signal Processing, vol. 2007, no. 1, pp. 141–141, 2007.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

[9] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 9, pp. 1627–1645, 2010.

[10] R. Brunelli, Template matching techniques in computer vision. Wiley Online Library, 2009.

[12] J. Lewis, "Fast normalized cross-correlation," in Vision interface, vol. 10, no. 1, 1995, pp. 120–123.

[13] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5–32, 2001.

[14] Sri Hari Nallamala, et al., "A Literature Survey on Data Mining Approach to Effectively Handle Cancer Treatment", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 729 – 732, March 2018.

[15] Sri Hari Nallamala, et.al., "An Appraisal on Recurrent Pattern Analysis Algorithm from the Net Monitor Records", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 542 – 545, March 2018.

[16] Sri Hari Nallamala, et.al, "Qualitative Metrics on Breast Cancer Diagnosis with Neuro Fuzzy Inference Systems", International Journal of Advanced Trends in Computer Science and Engineering, (IJATCSE), ISSN (ONLINE): 2278 – 3091, Vol. 8 No. 2, Page No: 259 – 264, March / April 2019.

[17] Sri Hari Nallamala, et.al, "Breast Cancer Detection using Machine Learning Way", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8, Issue-2S3, Page No: 1402 – 1405, July 2019.

[18] Sri Hari Nallamala, et.al, "Pedagogy and Reduction of K-nn Algorithm for Filtering Samples in the Breast Cancer Treatment", International Journal of Scientific and Technology Research, (IJSTR), ISSN: 2277-8616, Vol. 8, Issue 11, Page No: 2168 – 2173, November 2019.

[19] Kolla Bhanu Prakash, Sri Hari Nallamala, et al., "Accurate Hand Gesture Recognition using CNN and RNN Approaches" International Journal of Advanced Trends in Computer Science and Engineering, 9(3), May – June 2020, 3216 – 3222.

[20] Sri Hari Nallamala, et al., "A Review on 'Applications, Early Successes & Challenges of Big Data in Modern Healthcare Management'", Vol.83, May - June 2020 ISSN: 0193-4120 Page No. 11117 – 11121.

[21] Nallamala, S.H., et al., "A Brief Analysis of Collaborative and Content Based Filtering Algorithms used in Recommender Systems", IOP Conference Series: Materials Science and Engineering, 2020, 981(2), 022008.

[22] Nallamala, S.H., Mishra, P., Koneru, S.V., "Breast cancer detection using machine learning approaches", International Journal of Recent Technology and Engineering, 2019, 7(5), pp. 478–481.