Deep Learning

*Mr.Bhagwan K.Patil[1]
M.E Student
E and TC Engineering
D.Y. Patil College of Engg. And Tech, Kolhapur, Maharashtra, India

Prof.Dr.S.V. Sankpal[2]
Professor
Electronics Engineering
D.Y. Patil College of Engg. and Tech, Kolhapur, Maharashtra, India

Mr. Deepak Kamble[3]
Lecturer
E and TC Engineering
Sanjay Ghodawat Institute, Atigre, Maharashtra, India

Abstract - *We describe the problem of mixed sound event verification on a wireless sensor network for home automation systems. Sounds that are recognized by the system cause certain actions in home automation systems to occur. However, the system won't be able to accurately distinguish the initial sound if two noises occur simultaneously—a target source and another. This would result in erroneous replies. This research suggests a way to carry out sound-triggered automation in order to address such issues. The system uses Wireless Sensor Network (WSN) based techniques for sound separation and sound verification. We provide a Convolution Blind Source Separation (CBSS) system that estimates the source number for the sound separation phase using time-frequency clustering. The separated sound sources can be reconstructed by using the precise mixing matrix estimated by the proposed phase compensation approach. Mel Frequency Central Coefficients (MFCC) and Fisher scores—which are derived from the wavelet packet decomposition of signals—are utilized as support vector machine features during the verification phase. Finally, an automated service can be started by selecting a sound of interest based on the verification result. The experimental findings show that the suggested strategy for mixed sound verification in WSN-based home contexts is both feasible and reliable.*

***Keywords*: Wireless Sensor Network (WSN), Blind Source Separation (BSS), Support Vector Machine (SVM), Convolutional Neural Network (CNN).**

1. INTRODUCTION

The creation of inexpensive, low-power, multipurpose devices has been made feasible by advancements in sensor node technology. These are usually part of a Wireless Sensor Network (WSN), where a number of sensor nodes communicate with one another to gather data and keep an eye on the environment on a regular basis. Numerous applications, including as smart homes, environment sensing, and health monitoring, are already using WSNs. The main goal of this research is to construct a wireless sensor network that could be helpful for applications related to assisted living.

In this work, the sensor nodes collect audio streams, and the sink node processes the data to identify various environmental events. The two main contributions of this proposed work are the exploration of techniques that use different audio input representations are used in audio event detection using convolution neural networks (CNNs).Our everyday existence depends on our capacity to identify and collect information about physical quantities or processes, such as changes in temperature or pressure, that occur in the actual world. Sensing is a technique used to gather data on a physical quantity or process, including the occurrence of events (like changes in temperature or pressure).

A device that performs these kinds of sensing functions is called a sensor. The human body, for example, is equipped with sensors that detect optical information from the surroundings (eyes), auditory information from noises (ears), and odors (nose). A sensor is a device that transforms physical world parameters or events into signals that can be measured and analyzed. These are examples of remote sensors since they obtain data without having to come into contact with the object being monitored. The widespread use of sensors in devices, environments, and equipment has a net positive effect on civilization. They can shield valuable natural resources, stop catastrophic infrastructure failures, increase productivity, improve security, and enable cutting-edge applications like context-aware systems and smart home technology. The remarkable progress in semiconductor technologies led to the creation of microprocessors that have minimal dimensions and can process data at higher speeds. This makes it possible to design sensors, actuators, and controllers that are tiny, inexpensive, and low-power.

2. LITURATURE SURVEY

A strong sound recognition system for the environment intended for home automation. Sound classes that have been discovered can be used to trigger certain home automation services. Furthermore, human speech falls under the sound category, and this type of speech can be identified for identifying human intentions, just like in traditional home automation study. This system employs two primary methods to achieve this driven objective: analysis of independent components Mel-Frequency Cepstral Coefficients are used to identify sounds, and frame-based multiclass Support Vector Machines are used to improve subspace-based signals that are sensitive to noise ratio. Applications for home automation include energy management, lighting, heating, environmental control, security and safety, and appliance control [1]. The environment, people, or both can trigger these services. The suggested approach turned on home automation features by using auditory data. In the house, a variety of sound classes and human voice are common sources of acoustic information. There are known applications for human speech in home automation.

An E-Health system with wireless sensors. The design allows for round-the-clock continuous monitoring of senior citizens without interfering with their everyday routines or those of their caregivers. Both fixed and body (mobile) sensors are employed in the design system. Even if the home sensor network may monitor the health and living environment based on data provided by the wireless sensor network, it is not necessary to attach the wireless sensor board to the elderly person, and in many cases, they may not bring the sensor [2]. A mixed positioning algorithm is utilized to locate elderly people, which helps the system identify the person's actions and make judgments regarding their health.

Wireless sensor network-based home health care system. This device will close the communication gap between the patient and the doctor and enable early detection of unfavorable disease. This study includes five prototype designs for blood pressure monitoring, vital sign monitoring for firefighters, notifying partially or completely deaf individuals, and child monitoring [3].

Recent research has shown that deep learning techniques can be effectively applied to text and image data straight from raw sources. While not yet thoroughly investigated, this method has also been used with audio signals. In this work, we introduce a convolutional recurrent neural network that directly receives time-domain waveforms as input for the task of classifying urban noises. The convolutional recurrent neural network model combines the capabilities of recurrent neural networks for temporal aggregation of the extracts with convolutional neural networks for sound feature extraction[16].

Every element of a person's life is greatly impacted by sound. In a variety of domains, the study of sound classification has grown in interest recently. Sound is widely employed to construct automated systems in a number of businesses, from personal security to critical surveillance. Artificial intelligence plays a major role in speech recognition, which is growing more and more important as technology develops. Deep learning architectures are used in the development of sound classifier systems to solve the efficiency problems with traditional systems. This research presents a deep learning technique that uses generated spectrograms to categorize environmental noise. Deep learning systems are now capable of identifying the noises that people hear most frequently. Recently, environmental noise classification has been used to convolutional neural networks. An environmental noise detection convolutional neural network is trained using CNN spectrogram images[17].

Deep learning has several uses when it comes to audio signal classification. Speech, music, and background noise are just a few of the many audio signals it can recognize and classify. Deep learning models are highly accurate because they can be taught on massive datasets and can recognize intricate patterns in audio signals. The audio signal must first be represented in an appropriate way before deep learning can be used to classify the signal. Signal representation techniques including wavelet decomposition, linear predictive coding, Mel-frequency Cepstral coefficients, and spectrograms can be used for this. The audio signal can then be fed into a deep learning model once it has been appropriately represented. Different deep learning models can be applied to audio classification [18].

Numerous fields have made extensive use of sound classification. Deep learning technology is one of the most practical and efficient ways to classify sounds, in contrast to conventional signal-processing techniques. The classification performance is impacted by factors that limit the quality of the training dataset, such as resource and cost limitations, data imbalances, and problems with data annotation. In order to extract sound features, we propose a sound classification mechanism based on convolutional neural networks and employ Mel-Frequency Cepstral Coefficients (MFCCs) to transform sound data into spectrograms. CNN models can use spectrograms as input[19].

3. EXISTING VS PROPOSED BSS

A. *Existing System*

In existing system noise removal was applied after Blind Source Separation and SVM (Support Vector Machine) used for classification.

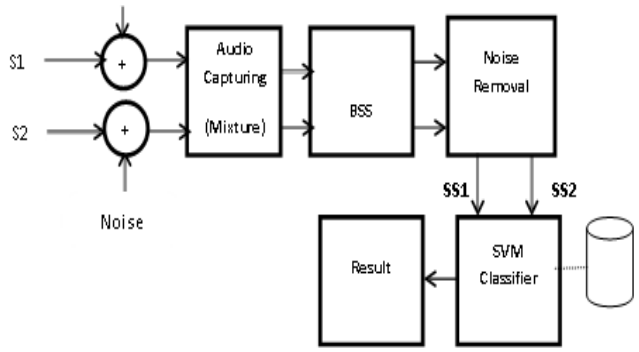

Fig. 1. Existing system block diagram

B. *Proposed System*

In proposed system noise removal was applied before Blind Source Separation , which makes more accurate signal separation and for classification Convolutional Neural Networks (CNN) is used.

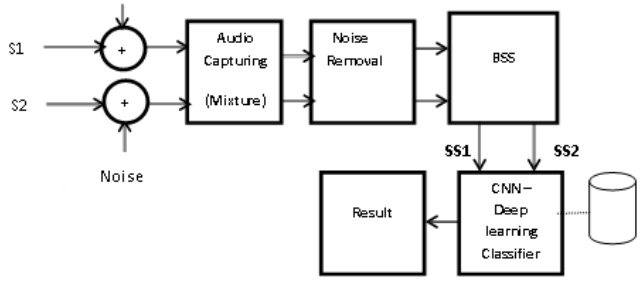

Fig. 2. Proposed system block diagram

4. PROBLEM STATEMENT

Verifying sound events is extremely difficult since several sound signals might blend together in real time, making it difficult to separate and categorize the signals. When a target source and another sound occur at the same time, the system's recognition performance will be compromised, leading to incorrect replies. Appropriate sound separation and sound verification techniques are needed to address this issue. Accurate sound categorization depends critically on the ability to separate sound from mixed signals with low levels of noise. Deep learning and classification are needed for more precise detection during the verification stage.

5. DATASET

1) Target sounds (sounds of interest) (clean)	
a) doorbell ringing	(20 audio files)
b) glass breaking	(20 audio files)
c) door knocking	(20 audio files)
2) Non target sounds	
a) cat meowing,	(30 audio files)
b) dog barking,	(30 audio files)

c) piano playing	(30 audio files)
d) human speech	(30 audio files)

The performance of our sound separation phase is the primary goal of our experiment. We then compare the verification results of mixed and separated sounds to demonstrate how the sound separation improves sound verification performance. Within our research, we employ three categories of signals as sounds of interest, also known as target noises: doorbell ringing, glass breaking, and hammering on doors. We additionally designate as non-target noises four undesirable sounds: dog barking, piano playing, cat meowing, and human talking.

6.METHODOLOGY

To improve separation and classification performance and accuracy, an automated technique has been devised. A convolutional neural network, or CNN, is used to categorize audio signals. Without the image going through any processing, a direct image (Mel Spectrum) is provided to the classifier in CNN, and probabilities are the result. The images in this project will be pre-processed before being sent to CNN. This project's primary goal is to categorize the target Sound. There are two stages:

Training phase:

Target Sound Signal Acquisition

Convert Mel Spectrum

Build training set

Create CNN model

Testing phase :

Signal Separation

Use CNN Model

Target Sound Classification.

Using characteristics taken from clean versions of these sound classes, we train our CNN classifier during the training phase. Additionally, the CNN classifier was trained using 30 clean audio clips gathered from the non-target classes and 20 clean audio files chosen from each target class. The sounds in our system have a duration of 10 seconds and a sample rate of 8 kHz. We use the mixed signal, which is a randomly chosen target sound combined with non-target sound, to test the system.

The sounds that are received at WSN are mixed, and since non-target sounds may be present, it will be difficult to distinguish the true target voice from the noise. The blind source separation (BSS) approach is used by the sink to separate mixed sounds upon receiving them. The signals are sent for testing and categorization after separation. A CNN model that has been trained using Mel Spectrum is used to classify the signals.

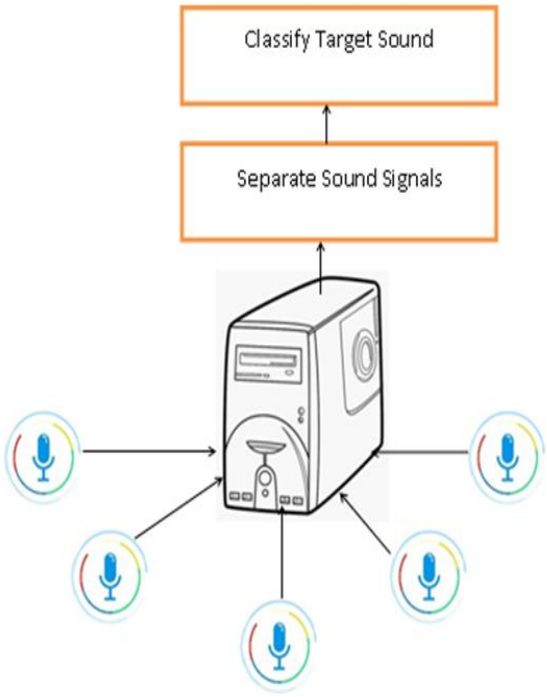

Fig. 3. System Architecture

A BSS Algorithm

Step 1. To estimate the necessary waveform-related variance for the channels, take a separate average of the observed signal and the reference signals of T trials:

$$x'(n) = \frac{1}{T}\sum_{t=1}^{T} x_t(n) \quad (1)$$

Step 2. To find deviations, subtract the averages from each trial set of data:

$$x''(n)=x(n)-x'(n) \quad (2)$$

Step 3. Utilizing linear least-square regression, determine the propagation factor C by treating the reference data as the independent variable and the observed signal data as the dependent variable:

$$X=C(X_{Ref}) \quad (3)$$

where

$$X=[\, x''(1)\ldots\ldots.\, x''(t)\ldots\ldots\ldots\ldots\, x''(T)]^{T}$$

$$x''(t)=[\, x''(1+N(t-1)),\ldots..\, x''(tN)]$$

Step 4. Subtract the reference data scaled by the propagation factor C from the observed signal data to make the correction:

$$x'''(n)= x(n)-C(x_{Ref}(n)) \quad (4)$$

B. Mel Spectrum generation

Identifying the parts of the audio signal that are useful for linguistic content detection and removing the remaining information, which includes background noise, emotion, mixed signals, etc., is known as feature extraction and is where any automatic speech recognition system begins. A popular feature extraction method in speech and audio processing is called MFCC. The spectrum properties of sound are represented by MFCCs in a fashion that is useful for a variety of machine learning applications, including music analysis and speech recognition.

To put it another way, MFCCs are a collection of coefficients that represent the form of a sound signal's power spectrum. The mel-scale is used to mimic how the human ear hears sound frequency after the raw audio

information has been converted into a frequency domain using a technique like the Discrete Fourier Transform (DFT).Ultimately, the mel-scaled spectrum is used to compute cepstral coefficients. Matlab provides function to plot Mel spectrum

S = melSpectrogram(audioIn , fs)

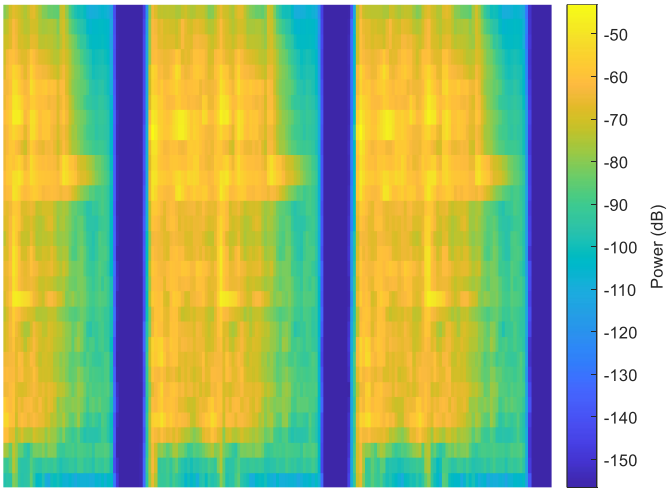

Fig. 4. Example of a Mel spectrogram

C. CNN classification

Five 1D convolution layers (Conv1D) make up the suggested CNN model. A number of additional layers come after each convolution layer, including batch normalization, MaxPooling, dropout, flattened, and dense layers. In our methodology, the sound data array and kernels are affected by the convolution layer. The extracted feature matrix is supplied into the first convolution layer, which uses the input sound files to produce a structural or detailed semantic feature map (local features). An array list with a size of 273 × 1 and a stride of one pixel makes up the input of the first Conv1D. There are 64 filters and five kernel sizes.

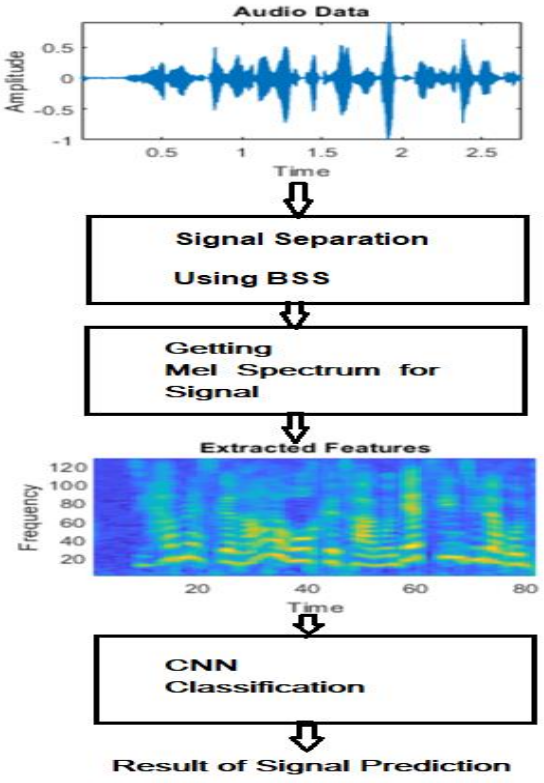

Fig. 5. CNN Classification

Batch normalization can mitigate the effects of unstable input data by scaling and shifting. The training phase then starts with a 20% attrition rate. The dropout layer helps to lessen the overfitting issue by removing input values that are smaller than the dropout rate. The dropout layer is followed by the max pooling layer. A one-dimensional max pooling layer with a size four pooling window has been selected. The Max Pooling layer helps with feature reduction by applying maximum filter activation at different places along the quantified windows to form a single output feature map.

D. DOA Estimation

The technique of estimating the direction of several electromagnetic waves or sources from the outputs of various receiving antennas that make up a sensor array is known as direction-of-arrival, or DOA estimation. With widespread applications in sonar, radar, wireless communications, and other domains, DOA estimation is a major problem in array signal processing.

7.RESULT

A.Performance Related Terminology and Definitions

- True Positive (TP): Accurately predicting the positive.
- False Positive (FP): An incorrect estimate for the positive.
- True Negative (TN): Accurately predicting the negative.
- False Negative (FN): An incorrect assessment for the negative category.
- **Accuracy: The** many types of accurate and inaccurate predictions have names that are peculiar to binary categorization. Thus, the following is the binary classification accuracy formula:
  Accuracy= (TP+TN) / (TP+TN+FP+FN) (5)
- **Sensitivity(SE):** True Positive Rate – TPR: The likelihood of a positive test result among signals classified as positive.

  SE = TP/(TP + FN) (6)
- **Specificity(SP) :** True Negative Rate – TNR: The likelihood of obtaining a negative test result from the signals classified as negative.

  SP = TN/(FP + TN) (7)
- **Precision:** What proportion of the model's predictions—the positive class—were accurate?

  Precision = TP/(TP+FP) (8)
- **Recall :** What proportion of predictions did the model accurately identify as the positive class when the ground truth was the positive class?

  Recall = TP/(TP+FN) (9)
- **False Positive Rate (FPR) :** The percentage of real negative examples for which the positive class was incorrectly predicted by the model. The false positive rate is computed using the formula below.:

  FPR=FP/(FP+TN) (10)
- **F1-score:** The F1 score is a machine learning evaluation statistic that measures a model's accuracy. It combines recall and precision ratings of a model.

$$F1\ Score = \frac{2 * precision * recall}{(precision + recall)} \quad (11)$$

B. Overall Performance with Mix wave - without separation

Table 1 shows performance analysis without separating the signal, i.e. directly using the noisy signals. Table clearly show the accuracy if very less.

TABLE I.

Accuracy	52.78 %
Error	41.67 %
Sensitivity	58.33 %
Specificity	79.17 %
Precision	74.92 %
Recall	58.33 %
False Positive Rate	20.83 %
F1 score	55.30 %

Even false positive rate is high and other metrics also not up to mark. The results clearly shows the signal separation is very essential. Figure 6 shows confusion matrix of CNN classification without separation i.e. mixed noisy signals. Confusion matrix shows class 1 (doorbell ringing) is reasonably good but class 2 (glass breaking) and class 3 (door knocking) prediction is very poor.

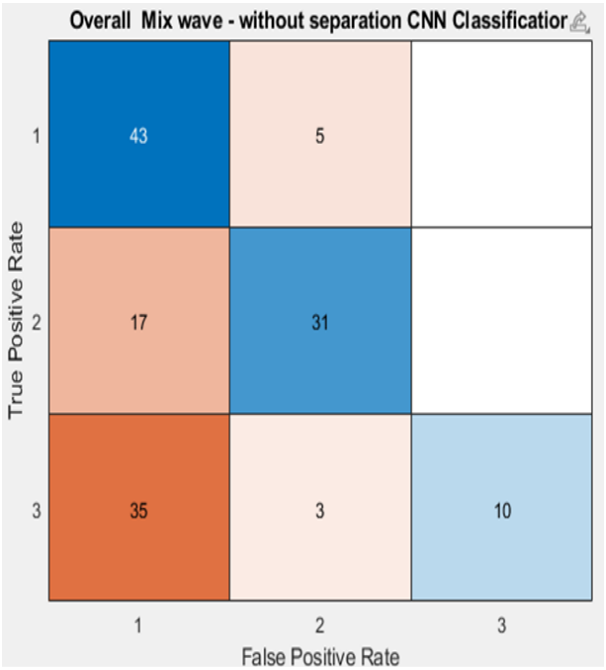

Fig. 6. CNN Confusion matrix of mixed signal

C.Overall Performance with separation (Separated wave)

Table 2 shows performance analysis if separating the signal , i.e. Separating target and non-target signals. Table clearly show the accuracy if excellent.

TABLE II.

Accuracy	93.06 %
Error	6.94 %
Sensitivity	93.06 %
Specificity	96.53 %
Precision	94.00 %
Recall	93.06 %
False Positive Rate	3.47 %
F1 score	93.13 %

False positive rate is very low and other metrics also up to mark The results clearly shows the classification after signal separation gives very good accuracy. Figure 7 shows confusion matrix of CNN classification after separation signals. Confusion matrix shows class 1 (doorbell ringing) , class 2 (glass breaking) , class 3 (door knocking) i.e. all classes is good and prediction is excellent.

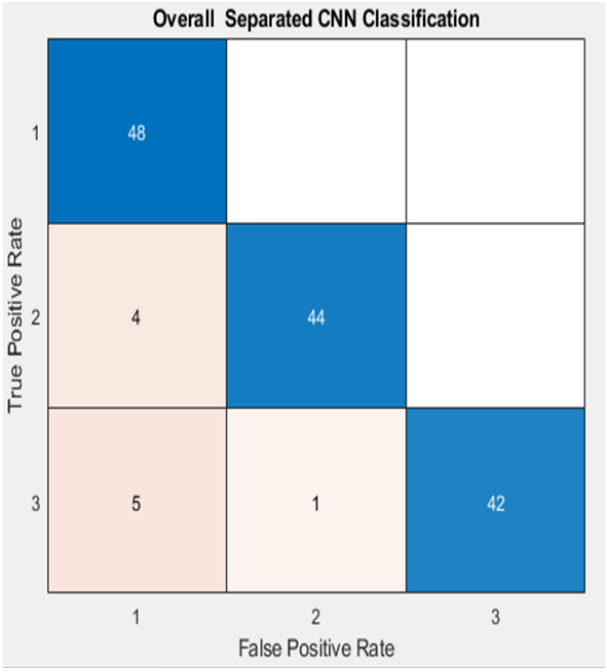

Fig. 7. CNN Confusion of Separating signal

Figure 6 makes clear visualization of the comparison existing method and proposed method by comparing accuracy and error and Figure 7 makes comparison of F1 Score.

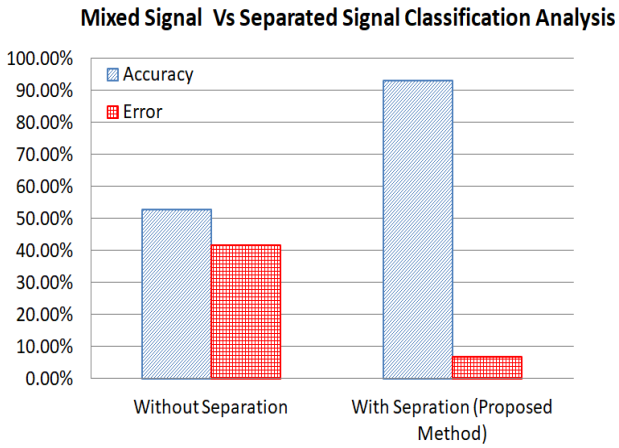

Fig. 8. Comparison between without separation and with separation

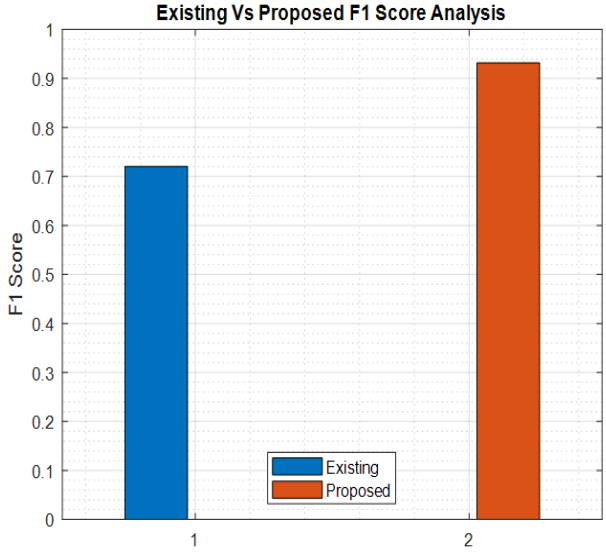

Fig. 9. Comparison of F1 Score

Table 3 , provides three DOA disparities (40, 90, and 160) for the average SIRs. The findings of the experiment suggest that source number estimation can have an accuracy rate of up to 100%.

TABLE III.

DOA Difference	40	90	160
SIR	18.16	18.16	26.63

Figure 10. shows as DOA angle increases above 90 SIR increases drastically

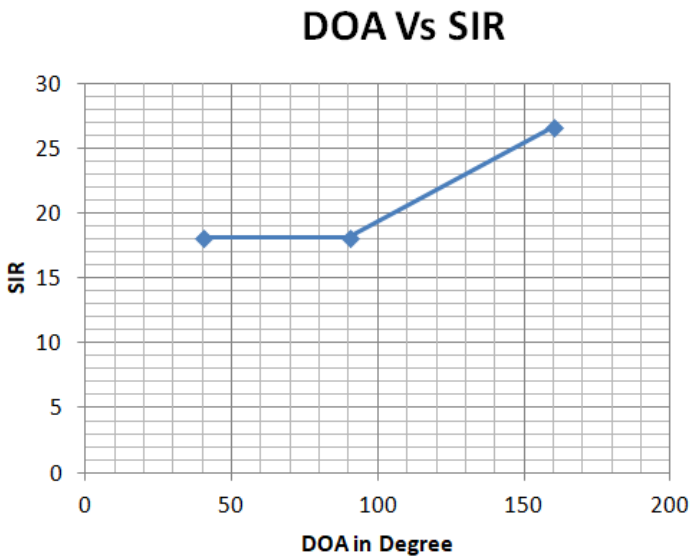

Fig. 10. DOA and SIR line graph

7.CONSLUSION

Generally speaking, feature extraction is the first step for speech recognition, In proposed work, the BSS (Blind Source Separation) is used for separating sound from mixes sound, The second step Mel Frequency Spectrum feature has been used for the speech classification system. The third step is training testing CNN which used for speech recognition system.

The training results show that a higher accuracy of 93.06% is obtained with a more accurate source separation. In the event when the sources are not divided, the accuracy is 52.78% or less. Therefore, before classifying, we suggest using BSS for sound separation in order to extract features from the environment.

8.REFERENCES

[1] *Chang-Hong Lin, Hsiao Ping Lee, Jyun-Hong Li, Chih-Wei Su, Yu-Hao Chin, Jhing-Fa Wang, Jia-Ching Wang: "Blind Signal Separation with Speech Enhancement",. EMC / Human Com 2013: 1049-1054 - 2013*

[2] *Hairong Yan, Hongwei Huo, "Wireless Sensor Network Based E-Health System – Implementation and Experimental Results", IEEE Transactions on Consumer Electronics, Vol. 56, No. 4, November 2010.*

[3] *Chris R. Baker, Kenneth Armijo, et al, "Wireless Sensor Networks for Home Health Care" , 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07) 0-7695-2847-3/07 $20.00 © 2007*

[4] *Michel Vacher, Dan Istrate, et al , The SWEET-HOME Project: Audio Technology in Smart Homes to improve Well-being and Reliance , 33rd Annual International Conference of the IEEE EMBS Boston, Massachusetts USA, August 30 - September 3, 2011.*

[5] *Jianfeng Chen, Alvin Harvey Kam, et al "Bathroom Activity Monitoring Based on Sound" , PERVASIVE 2005, LNCS 3468, pp. 47 – 61, 2005. © Springer-Verlag Berlin Heidelberg 2005.*

[6] *Vehbi C. Gungor , Bin Lu , Gerhard P. Hancke ,"Opportunities and Challenges of Wireless Sensor Networks in Smart Grid" , 2010 IEEE Transactions on Industrial Electronics*

[7] *Shreya Sose , Swapnil Mali , S.P. Mahajan , "Sound Source Separation Using Neural Network" , 10th ICCCNT 2019 July 6-8, 2019, IIT - Kanpur Kanpur, India.*

[8] *Mohammed Safwan, Deepak S G, "A COMPARATIVE STUDY OF BLIND SOURCE SEPARATION BASED ON PCA AND ICA" , © 2022 IJCRT | Volume 10, Issue 6 June 2022 | ISSN: 2320-2882.*

[9] *M. Hassan Tanveer , Hongxiao Zhu et al , "Mel-spectrogram and Deep CNN Based Representation Learning from Bio-Sonar Implementation on UAVs" , 2021 International Conference on Computer, Control and Robotics*

[10] *Yang Li , Ji Maorong, et al. , "Design of Home Automation System based on ZigBee Wireless Sensor Network" , The 1st International Conference on Information Science and Engineering (ICISE2009).*

[11] *Wang Huiyong, Wang Jingyang, Huang Min , "Building a Smart Home System with WSN and Service Robot" , 2013 Fifth Conference on Measuring Technology and Mechatronics Automation.*

[12] *J. Brito, T. Gomes, et al , An Intelligent Home Automation Control System Based on A Novel Heat Pump and Wireless Sensor Networks , 978-1-4799-2399-1/14/$31.00 ©2014 IEEE*

[13] *Nikˇsa Skeledˇzija, Josip ´ Cesi´ c, et al , "Smart Home Automation System for Energy Efficient Housing" , MIPRO 2014, 26-30 May 2014, Opatija, Croatia*

[14] *Rasika S. Ransing , Manita Rajput , "Smart Home for Elderly Care, based on Wireless Sensor Network" , 2015 International Conference on Nascent Technologies in the Engineering Field (ICNTE-2015).*

[15] *Vikram.N1, Harish K.S2, et al , A Low Cost Home Automation System Using Wi-Fi Based Wireless Sensor Network Incorporating Internet of Things(IoT) , 2017 IEEE 7th International Advance Computing Conference.*

[16] *Waveforms Jonghee Sang, Soomyung Park et.al. Convolutional Recurrent Neural Networks for Urban Sound Classification using Raw, european signal processing conference , 2018*

[17] *Abhay Chaturvedi; Suman Avdhesh Yadav; et.al Classification of Sound using Convolutional Neural Networks, Conferences >2022 5th*

[18] *Khalid Zaman; Melike Sah; Cem Direkoglu; Masashi Unoki , A Survey of Audio Classification Using Deep Learning, Journals & Magazines >IEEE Access >Volume: 11*

[19] *Hung-Chi Chu, et.al. A CNN Sound Classification Mechanism Using Data Augmentation, Published online 2023 Aug 5. doi: 10.3390/s23156972*