ARTIFICIAL NEURAL NETWORK-BASED SPEECH RECOGNITION USING DWT ANALYSIS APPLIED ON ISOLATED WORDS FROM ORIENTAL LANGUAGES

***Bacha Rehmam*[1], *Zahid Halim*[2], *Ghulam Abbas*[3], *Tufail Muhammad*[4]**

[1, 2, 3, 4] Faculty of Computer Science & Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, Pakistan.

Email: [1] bacha @giki.edu.pk, [2]zahid.halim@giki.edu.pk, [3]abbasg@giki.edu.pk, [4]tufail @giki.edu.pk

ABSTRACT

*Speech recognition is an emerging research area having its focus on human computer interactions (HCI) and expert systems. Analyzing speech signals are often tricky for processing, due to the non-stationary nature of audio signals. The work in this paper presents a system for speaker independent speech recognition, which is tested on isolated words from three oriental languages, i.e., Urdu,*Persian, and Pashto. *The proposed approach combines discrete wavelet transform (DWT) and feed-forward artificial neural network (FFANN) for the purpose of speech recognition. DWT is used for feature extraction and the FFANN is utilized for the classification purpose. The task of isolated word recognition is accomplished with speech signal capturing, creating a code bank of speech samples, and then by applying pre-processing techniques.For classifying a wave sample, four layered FFANN model is used with resilient back-propagation (Rprop). The proposed system yields high accuracy for two and five classes.For db-8 level-5 DWT filter 98.40%, 95.73%, and 95.20% accuracy rate is achieved with 10, 15, and 20 classes, respectively. Haar level-5 DWT filter shows 97.20%, 94.40%, and 91% accuracy ratefor 10, 15, and 20 classes, respectively. The proposed system is also compared with a baseline method where it shows better performance. The proposed system can be utilized as a communication interface to computing and mobile devices for low literacy regions.*

Keywords: Speech recognition, Artificialneural networks, Discrete wavelet transform, Feature extraction.

1.0 INTRODUCTION

Speech is a normal way of communication for humans. People learn all the necessary and relevant skills in the early childhood, with or without instructions, and rely on communication using speech. The key factor behind this process is the human speaking tract and articulation abilities. These are biological organs with properties of non-linearity, whose operations are not merely under conscious management, but are affected by factors varying from gender, childhood,and state of emotion. As a consequence, the vocalizations can change broadly in terms of their pronunciation, accents, articulation, nasality, roughness, volume, pitch, and speed. Furthermore, in the time of transmission, peoples' irregular patterns of speech can be more distorted by the background noise and echoes, as well as electrical properties or characteristics in case telephones or other electronic devices are used. All these sources of variability make speech recognition a complicated processas compared to the generation of speech[1].

Speech recognition is the processthat converts an acoustic signal (acquired with the help of a microphone or telephone) to a collection of text or words. The recognized words can be ultimatelyused for applications, such as: commands and control, data entry,assistive technology [24], and document preparation. These can also be used as the input to other linguistic processing models in order to achieve speech understanding. Recognition is generally consideredto be more complex when vocabularies are large or have numerous similar sounding words.

This work mainly focuses on Urdu speech recognition, as the proposed methodology is tested for the speech samples spoken inUrdu. "Urdu" is a Turkish word which means horde ("لشکر"). Urdu is the national language of Pakistan and a major secondary language in the Indian subcontinent. Urdu fundamentally relates to oriental language, which is a sub-branch of Indo-Aryan, developed with the influence ofPersian, Turkish, and Arabic [2]. Urdu script is written from right to left similar to Arabic, Pashto, and Persian. There is an immense need for developing a speech recognition application for Urdu as a high percentage of Urdu speaking regions have relatively low literacy rate and, as a result, speech based interface will be very helpful in providing them with access to information [3]. This paper presents a novel speech recognition application for the Urdu language. In addition to the language of focus, i.e., Urdu, we have also performed experiments on two other oriental languages, including: Pashto and Persian. The choice of these languages is based on their local and global influences. Urdu, Persian, and Pashto are spoken in multiple countries, including: Pakistan, India, Afghanistan, and Iran.

Some speech recognition systems require a speaker to provide samples of speech before using it. These kinds of systems are custom built for a singleuser, and are, therefore, not commercially feasible. Whereas,for other types of speech recognition systems,speaker enrolment is not required in advance.Such kind of generic speech recognition systems are more likely to become familiar with the specific speakers they are trained on, resulting in lower recognition accuracy thanthe speaker dependent systems. For the isolated-word speech recognition system, speakers are required to pause briefly between spoken words. Conversely, in a continuous speech recognition system,there is no such restriction. A speech recognition system can be trained to recognize whole words. This kind of training methodology is useful in applications resembling voice command systems, in which the speech system requiresto merely recognize a small set of words. This approach is called keyword based training [4]. The method presented in this paper is based on isolated and keywords.

The objective of this paper is to use computational intelligence for speech recognition applied on isolated words of the oriental languages. Motivated by the low literacy rates of many regions of the oriental origin [25], the proposed approach in this paper bridges the gap between the cutting-edge technology and the regional languages. The proposed system aims to provide a communication interface to illiterate people so that they can benefit from computing technologies using voice commands in their native language. Most of the existing commercially available products such as Google now [26], Microsoft's Cortana [27], and Apple's Siri [28], areavailable for English. In addition to these, other recent solution for the oriental languages like, [30], [31], and [32]is based on text editing which again requires a well literate user.

In this paper,discrete wavelet transform (DWT) analysis andfeed forward artificial neural network (FFANN) with back-propagation isinvestigated and applied,with the aim of producing better results for speaker-independent speech recognition system tested on isolated words under general environment. For speech recognition, DWT is used with db-8[5], haar, and sym-8[6- 7] filters after applying pre-processing techniques on the speech sample to extract the energy in the wavelet coefficient as well as the wavelet sub-band's coefficients. The feature vector acquired from the energy calculated from each sub-band of the wavelet coefficientis tested by FFANN with two hidden layers.The main focus of this work is to develop a system that can be integrated with any speech recognition application for oriental languages. There are many applications of the proposed system, including theuser interfaces for farmers to know the yield details, in-car systems, assistive technology for disabled [24], computer gaming [44], and communication interface for low literacy regions [25].

The rest of the paper is organized as follows. Section 2 discusses the related work. Section 3 lists the pre-processing techniques applied in this paper. Section 4 describes the basic concepts of DWT analysis, the reason for choosing DWT and finally the proposed feature extraction method. Section 5 discusses the general concept of ANN and the proposed model. Section 6 lists theexperiments. Section 7 presents the experiments on other oriental languages and comparison with the baseline. Finally, Section 8 concludes the paper.

2.0 RELATED WORK

Much work has been done in speech recognition[8-9- 10].In order to recognize speech,a computational intelligence-based solution usually makes use of a classifier. Three most commonly used classifiers are: dynamic time warping (DTW), hidden Markov model (HMM), and artificial neural networks (ANN) [11].Different researchers have used these techniques in various ways and with diverse constraints to obtainrobust results. In [1], a combination of ANN with HMM is presented withspatio-temporal ANN used for a small set of speech samples. In [12], a wavelet-based features vector is tested with DTW. The solution in [12] is focused towards a fault diagnosis system for internal combustion engines. In [12], there is another approach for speech recognition using wavelet transformation and ANN as the final classifier for a speaker dependent ASR. The study develops a wavelet packet adaptive network-based fuzzy inference system (WPANFIS). An accuracy of 92% for the sample speech signals is reported. There exists detailed methodologiesin[13- 14- 15- 16] for Urdu speech recognition by using Short-Time Fourier Transformation (STFT) with ANN as a final classifier. STFT is the analysis technique based on the constant window size when analyzing the speech sample. The proposed technique is based on the wavelet analysis which can analyze a speech sample using a variable window size based on the frequency of the speech segment.

A speech recognition support system named BizVoice is proposed in [33] that accessspeech during a session and converts the voice data to text data with a speech recognition technology. The effectiveness of the system is evaluated during a classroom session at Aoyama GakuinUniversity. A system based on a combination of the deep bidirectional LSTM recurrent neural network architecture and the Connectionist Temporal classification objective function for speech recognition is presented in [34]. The approach directly transcribes audio data with text. The error rate of 6.7% is reported. The work demonstrates character-level speech transcription by a recurrent neural network. An interesting approach for speech recognition using convolutional neural networks (CNNs) is presented in [35]. Results show reduction of error by a margin of 6-10% due to CNNs. Speech in [35] is analyzed using a 25-ms Hamming window with a fixed 10-ms frame rate. Data samples are normalized to a zero mean and unit variance. An approximate method for converting a feedforward NNLM into a back-off n-gram language model is proposed in [36]. The model is then used for speech recognition. The approach can be applied to non-back-off language model for efficient decoding.The work in [37] utilizes recurrent deep neural networks for speech recognition. The approach is evaluated on 2nd CHiME challenge (track 2) and Aurora-4 tasks showing an improvement of 7% and 4%, respectively.An accelerated FPGA implementation of Lithuanian isolated word recognition system is presented in [38]. For the purpose of testing the approach in [38], two databases are utilized. The first database has 100 words, each pronounced tentimes. The second database consists of ten words each pronounced ten times by ten speakers.The work in [39] presents a (Dynamic Time Warping) DTW-based approach for speech recognition in an intelligent electric wheelchair. Key features are obtained using Mel Frequency Cepstral Coefficients (MFCCs). The voice signals are then transformed to various commands.. An approach based on linear discriminant analysis is presented in [40] to recognize isolated words of Urdu. For every item in the database, 52 Mel Frequency Cepstral Coefficients are extracted for the classification purpose. Audio samples of seven speakers have been used in the system. The percentage error of less than 33% is reported for majority of samples. A comparative analysis of DWT-based features and Mel Frequency Cepstral Coefficients (MFCC) for speech recognition of Urdu language is presented in [41]. The database of words is selected from the most frequently used Urdu words. A system using an open source speech recognition framework called Sphinx4, has been presented in [42] for speech recognition. The speech recognition targets Urdu language with a vocabulary of 52 isolated words.

As evident from the aforementioned literature survey, there is limited work available for speech recognition using oriental languages. There is a large amount of work available on recognizing English words. However, areas with low literacy rates require speech recognition systems based on their local languages. This is one of the major shortcomings in the previous work. The literature review also reveilles ANN and DWT being used more prominently for speech recognition. The approaches proposed in [36] produce promising results, but the literature identifies a long training time and limitations on the number of context words taken into consideration as their limitation. Similarly, the proposal in [39] suffers from the issues of local optimum.Based on these issues, the work presented in this paper deals with presenting a speech recognition approach using ANN and DWT for isolated words from the oriental languages.

3.0 PRE-PROCESSING

In pre-processing, speech related information is extracted from the speech signal. Pre-processing includes techniques like removing noise from the speech samples, pre-emphasis, silence removal from the speech samples, de-emphasis, and normalization of the speech samples. Another important pre-processing technique applied is to divide speech samples into frames or different parts. The pre-processing techniques applied on the speech samples in this work are illustrated in Fig. 1. We start by capturing a speech sample and apply pre-emphases. Later the noise and silence portions of the signal are separated. This is followed by the normalization and framing of the recorded samples. These steps contribute towards the pre-processing of the audio sample recorded for the speech recognition purpose.

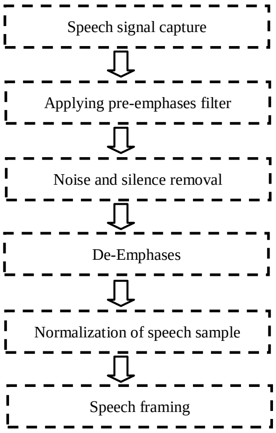

Speech signal capture
Applying pre-emphases filter
Noise and silence removal
De-Emphases
Normalization of speech sample
Speech framing

Fig.1: Pre-processing techniques applied

3.1 Speech signal capturing

In signal processing, sampling is the drop-off of a given continuous signal to a discrete signal. A sample represents a value or multiple values at any point in a particular time with respect to space. The sampling frequency (sometimes called sampling rate), denoted by F_s, can be defined as the quantity of samples per second.The equation of sampling rate is $f_s = 1/T$, where T is the time period of the signal. The sampling frequency is measured in Hertz, which is the number of samples per second.The input Urdu word spoken by a user is captured and recorded through a microphone. The captured speech signal is then stored asa waveform audio file format. A sampling rate of 16000 Hz is used for all speech samples. A total of 200 samples isrecorded for 20 Urdu words from tenspeakers under general environment. Table 1 illustrates the speech database recorded for the proposed system.

Table 1. Speech database

Sr. No.	Pronounce as	Equal to English word	No. of recorded samples
1	Aik	One	10
2	Do	Two	10
3	Teen	Three	10
4	Chaar	Four	10
5	Paanch	Five	10
6	Che	Six	10
7	Saat	Seven	10
8	Aath	Eight	10
9	Nau	Nine	10
10	Das	Ten	10
11	Gyara	Eleven	10
12	Baara	Twelve	10
13	Taira	Thirteen	10
14	Choda	Fourteen	10
15	Pandra	Fifteen	10
16	Sola	Sixteen	10
17	Satra	Seventeen	10
18	Athara	Eighteen	10
19	Unnees	Nineteen	10
20	Bees	Twenty	10

3.2 Pre-Emphasis

Human voice typically has the trait that the signal components possessing lower frequencies encompass higher amplitudes[17], where the higher frequencies encompass comparatively smaller amplitude. In this phenomenon, the higher frequencies (which have lower amplitude) have a lot of chances to be eliminated from the signal in the noise removal process as a noise. Hence there is a need to emphasize the high frequency components of a speech signal. For this purpose, pre-emphasis is performed, because it raises the signal-to-noise ratio (SNR). For pre-emphasis, the following procedure has been adopted.

Let *x*be an input sound signal and*px* is the resultant signal obtained after pre-emphasis, as in Eq. (1).

$$px\ (n) = x\ (n) - 0.97\ (n-1), \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (1)$$

where*n* is the number of samples in the entire speech signal. The coefficient 0.97 in Eq. (1) depends upon the order of finite-impulse response (*FIR*) filter used. After this, the speech signal is sent to the noise-removal procedure.

3.3 De-noising and Voice Activated Region Extraction

This process is used to cut off the recorded spoken speech sample from its neighbouring noise and silence. Induction of distortion in a speech signal fromthe surrounded environment is called noise[18]. The gap a speaker avails during the recording of a speech sample is indicated as a silence zone. There is a need ofa de-noised speech sample with the existence of regionshavingthe voice in it. The unvoiced portion of a speech signal is supposed to be an aperiodic or random waveform, while the voiced portion of a speech signal has a periodic waveform.

Normally,the first 200 milliseconds of speech recording match as unvoiced region. This is because of the time taken by speaker to speak while starting the recording. As stated by the Central Limit Theorem, all the unvoiced regions of a particular speech signal will pursue a normal distribution with a mean and standard deviation. We can assume *mu* as the mean and *sigma* as the standard deviation of the first 200m-sec of that particular speech signal. For a normal distribution under the 68-95-99.7 rule [19], 99% of the data in a speech signal lies at a distance of 3 × sigma from the mean; 95% of the data in the speech signal lies at a distance of 2 × sigma from the mean;and 68% of the data in the speech signal lies at distance of 1 × sigma from the mean; as represented by Eq. (2), Eq. (3), and Eq. (4) respectively.

$$P(\ ABS\ (s - mu) < 3 \times sigma) = 0.99 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (2)$$
$$P\ (\ ABS\ (s - mu) < 2 \times sigma) = 0.95 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (3)$$
$$P\ (\ ABS\ (s - mu) < 1 \times sigma) = 0.68 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (4)$$

If the distance of *s* from the mean *mu* is smaller than 3, we can declare with 99% confidence that this particular speech sample is noise[20]. This can be represented mathematically as:

$$D = ABS\ (s - mu) \div sigma \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (5)$$

Next, the speech samples which are recognized as noise are marked zero, whereas those recognized as voice are marked one. The resulting array of both zeroes and ones is divided into (non-overlapping) windows of 10 m-sec of length. Either a one or zero is assigned to all the samples contained in a particular window on the basis of the majority of one’s or zeroes that exist in a particular window. Fig.2 shows the original speech sample for the Urdu word “paanch”, while Fig.3 shows the entire voiced region extracted by selecting only those windows which are marked as ones.

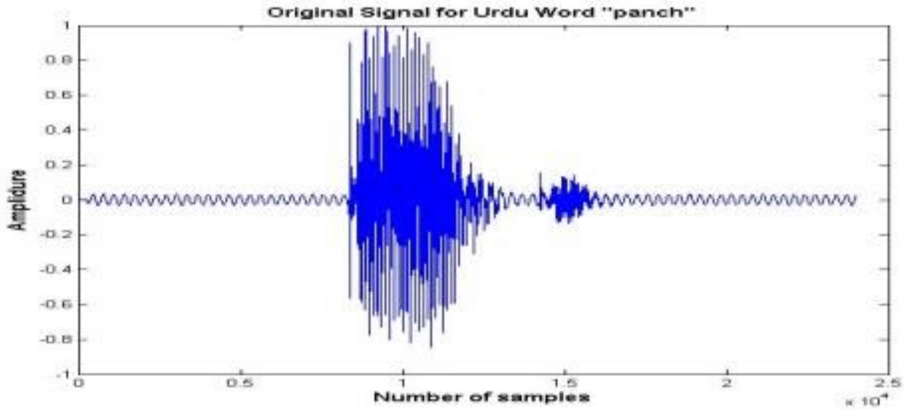


Fig.2: Original speech sample for "paanch"

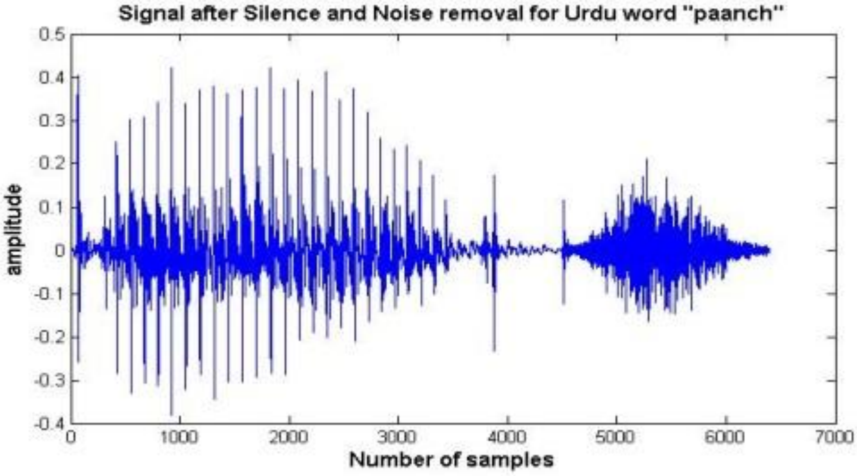


Fig.3: Signal after noise and silence removal

3.4 De-emphasis

De-emphasis is exactly the opposite process of pre-emphasis. In this process, a speech sample is converted back to its original amplitude and frequency after removing the noise and silence. The whole process of de-emphasis together with pre-emphasis is called emphasis. The equation for de-emphasis is given as:

$$dx\,(n) = px\,(n) + 0.97\,(n-1), \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (6)$$

where$n$ is the number of samples contained in the entire speech signal.

3.5 Normalization

A speech signal is normalized to make sure that the volume of a speaker during the recording of speech signal does not have an effect on the analysis as follows.

$$x= ((x - Mean\,(x)))/(MAX\,(ABS\,((x - Mean\,(x)))), \ldots\ldots\ldots\ldots\ldots\ldots(7)$$

where$x$ is a speech signal acquired after de-emphasis.Fig. 4 shows the normalized speech signal of the Urdu word “paanch”.

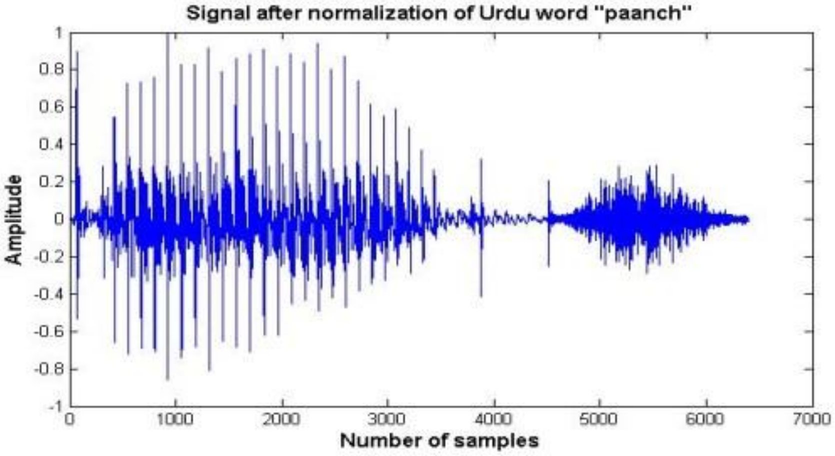


Fig.4: Signal after Normalization

3.6 Speech framing

After normalization, the speech signal is divided into three sets of identical length, with 33% overlap. This phenomenon means that the speech signal is separated into 3 × groups of equal duration, such that all the groups share 33% values in common. Speech framing is done because speech signals are not stationary. Suppose that the length of a sound signal is (2×L+1), then the first group contains values from 1 to (9 × L ÷ 7). The second group contains values from (6 × L ÷ 7) to (15 × L ÷ 7). Finally the third group contains values from (12 × L ÷ 7) to (2×L+1). Therefore, each group has an estimated length of (9 × L ÷ 7).Fig. 5 shows the three frames of the resultant speech signal after normalization of the Urdu word “paanch”.

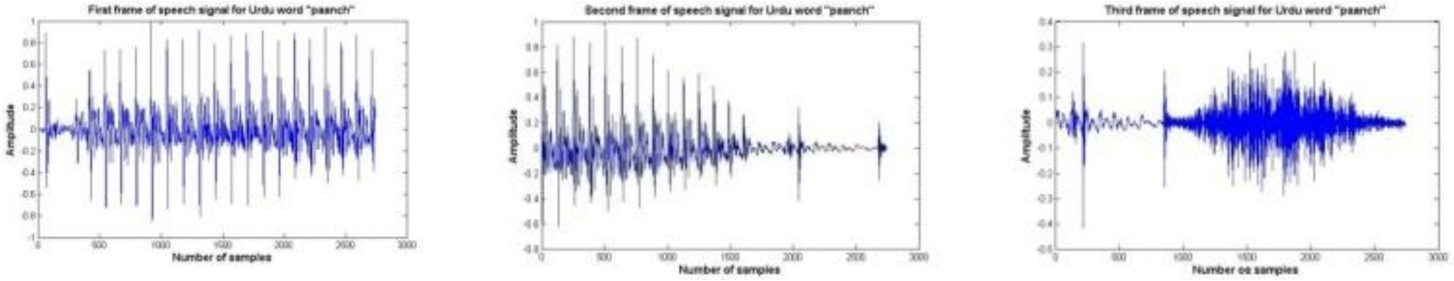


Fig.5: Speech signal after performing framing

During the speech recognition process, information for every speech sample (both training and testing set) is stored in a codebook or a vector array of$m$×n dimension,where m is the length of the speech signal and n is the number of speech samples fed to the system for extracting the features from it.

4.0 FEATURE EXTRACTION

Feature extraction plays a key rolein the Automatic Speech Recognition (ASR) system, and is possibly the most significant component of developing an intelligent system on the basis of speech or speaker recognition. Features extracted have an effect on the performance of a classifier[12]. A feature extracting mechanism should be able to reduce the pattern vector (the original waveform) to a much lower dimension, which may have most of the critical information that was present in the original vector.

A feature can be said to be a minimal unit, which is able to distinguish maximally close up units. The extracted feature vector should have the following characteristics: (a) the dimension should be small, (b) it should be stable for a long time interval, (c) it should be easy to compute from the samples of input speech, (d) it is able to alert extensively from one class to another, (e) it is not sensitive to the unrelated deviation in the speech input, and (f) there should be no correlation with the other features.

4.1. Discrete wavelet transform (DWT)

The next step after the processing of speech is to use the DWT to change the input speech signal to a frequency domain from the time domain. Some significant features can only be analyzed in the frequency domain, thus, it is common to examine the speech signal in the frequency domain. To enlighten what is the discrete wavelet transform, we have to first clarify the concept of continuous wavelet transform (CWT). This is because a DWT is devised by a CWT. A CWT can be defined as the sum over the entire signal's time multiplied by its scaled and shifted versions of the wavelet function (Ψ)[21]:

$$C\ (s,p) = \int_{-\infty}^{\infty} f(t)\Psi(s,p,t)dt, \ldots.. (8)$$

where*s* is the scale and *p* is the position of the wavelet. In Fig. 6, a CWT breaks one signal into sub-wavelets of diverse scales and positions.

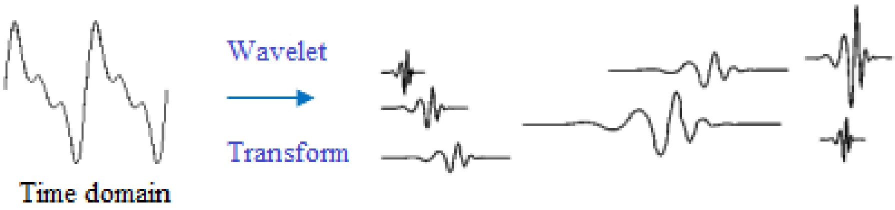


Fig.6: Transformation of a signal into scale and position

It ismore efficient if only a subset of scales and positions is selected. The technique in which the scales and positions are selected on powers of two bases is called discrete wavelet transform (DWT). Wavelet transform (WT) can be applied to signals that are not stationary. It concentrates into minute parts of the signal which can be well thought-out as stationary. WT has a changeable size window dissimilar to the constant size window used in short time Fourier transformation (STFT). Wavelet transform gives information about the type of frequencies' band in a given period of time. There are two main approaches using wavelet for speech decomposition, i.e., DWT and the wavelet packet decomposition (WPD) [11]. For this work DWT is used.

In DWT, the transform of a particular signal is just a different variety of representing the signal. It does not change the content of information of the signal. For numerous signals, the low-frequency part holds the most significant portion because it provides an identity to a signal. In wavelet analysis, approximations and details are often considered. The approximations are the low-frequency, high-scale components of the signal. Where, the details are in the high frequency components, low-scale. The DWT can be defined by the following Eq.(9):

$$W(j,k) = \sum\sum x(k)2^{-\frac{j}{2}}\Psi(2^{-j}n - k), \ldots\ldots.. (9)$$

where Ψ(t) is called the mother wavelet which is a time function with finite energy and fast decay. The time domain signal's consecutive high-pass and low-pass filtering is defined by the following equations:

$$\mathrm{y_{low}}\ [\mathrm{n}] = \sum_{\mathrm{k}=-\infty}^{\infty} \mathrm{x}\ [\mathrm{k}].\,\mathrm{g}\ [2\mathrm{n}-\mathrm{k}], \ldots\ldots\ldots. (10)$$
$$\mathrm{y_{high}}\ [\mathrm{n}] = \sum_{\mathrm{k}=-\infty}^{\infty} \mathrm{x}[\mathrm{k}].\,\mathrm{h}\ [2\mathrm{n}-\mathrm{k}], \ldots\ldots... (11)$$

where*g* represents the approximate coefficient and *h* represents the detailed coefficients. In DWT, by passing the preceding approximation coefficients, each level is calculated through a high and low pass filters.

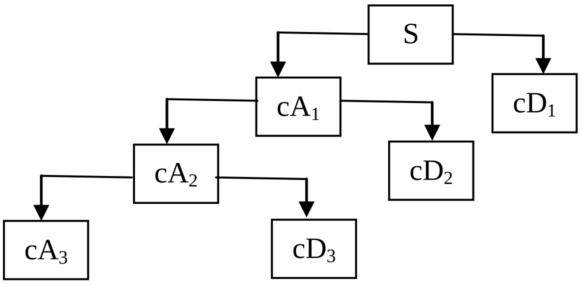


Fig.7: Decomposition tree

There are many types or families of wavelets (see Fig. 8), which are corresponded beside the shape of the desired signal whose wavelet transform is required. The wavelet is selected on the basis of its similarity in the shape [21] to the desired physical characteristics of the signal.

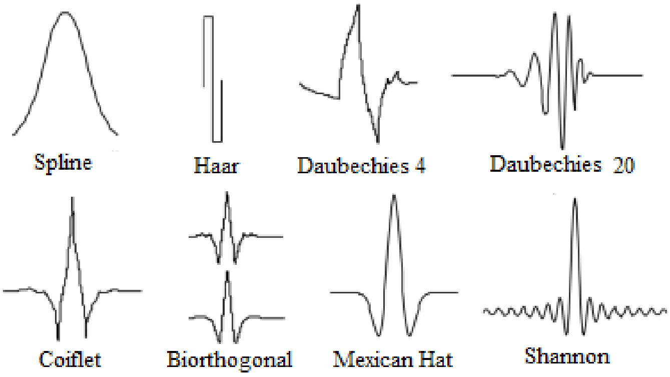


Fig.8: Some Types of wavelet families

4.2. Proposed technique for feature vector extraction

As discussed in section 3.6, the speech signal is divided into three sets of identical length, with 33% overlap. For feature extraction, each set is independently decomposed by using DWT, to determine the efficiency of these wavelet families for speech recognition. Three types or families are used for decomposing the speech signal,that is,haar, db-8, and sym-8 [5-6- 7]. This is because there is a need. Each frame of the original signal is decomposed by DWT [21] using haar, db-8, and sym-8 mother wavelet up to level 5. The number of levels for all the applied wavelet families isthe same because the same number of feature vectors for all the cases is required. DWT decomposition facilitates analysing a speech signal at dissimilar frequency bands. At every level of DWT, there exist two types of filters [21] (low-pass and high-pass filter). The decomposition of a signal *S* into these filters is shown in Fig.9.

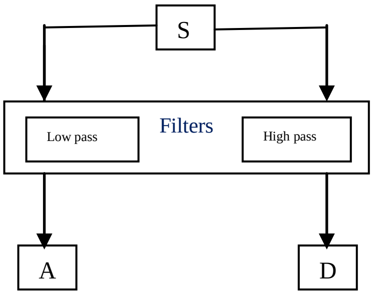


Fig.9: Signal *S* passes through low and high pass filters to produce signal *A* and *D*

If the frequency range of an input speech signal is from 0 to *f*, then the frequencies above *(f÷2)*are cut off by the low-pass filter, while the frequencies below *(f÷2)*are cut off by the high-pass filter. Filtering a speech signal is equivalent to the convolution mathematical operation of the signal [22- 23] by the filter's impulse response (see Eq. (10) and Eq. (11)). In DWT decomposition case, these filters correspond to the mother wavelet coefficients.

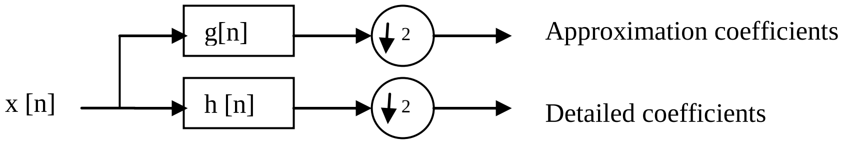


Fig.10: Signal *x[n]* is passed through both low and high pass filters and it is down sampled by 2

Every DWT decomposition is trailed by sub-sampling of 2 (see Fig. 10). After 5 levels of decomposition, 6 DWT coefficients are obtained for every part out of three parts for the whole speech signal. Therefore, a sum of 18 DWT coefficients is obtained for the whole signal. After this, the energy percentage of every part of the speech signal and also the energy in every DWT coefficient (sub-bands) is calculated. In other words, the percentage energy of the entire speech frame as well as the energy in every sub-band of the frame (as DWT coefficient) is calculated. At the end, an array of 18 values of energy coefficients for every input speech signal is acquired by merging the energy values of all three speech parts. Hence, 18 features vector for every speech sample are obtained.

A hybrid of DWT and FFANN is used in this work. The DWT is used to change the input speech signal to the frequency domain from the time domain. The major reason for using DWT for feature extraction is its capability to capture both frequency and location information as compared to the Fourier transform. On the other hand, the FFANN has the advantage of rapid training and the flexibility of addition/removal of training samples.

5.0 CLASSIFICATION MODEL

In speech recognition, there are three common methods for classification: Hidden Markov Model (HMM), DTW[12], and ANNs. Currently ANNs isemployed in several researches because of its parallel distributed processing capabilities, error stability, distributed memories, and distinguishing ability of pattern learning [11].

In the process of selecting the best classifier for the proposed system, we developed and tested the ANNs in various configurations, especially the size and the number of hidden layers. Table 2 shows some of the accuracyresults for 20 classes case applied on db-8 level 5 features. A maximum accuracy of 95.2% was obtained with two hidden layered ANN model (18, 20, 20, 1). Therefore,this is the selected network model for the proposed system.

Table 2. Accuracy rates of different ANN models

Neurons	Accuracy
18,5,5	68%
18,10,10	85.35%
18,20,20	88.62%
18,5,5,1	74.33%
18,10,10,20	82%
18,10,10,1	91.25%
18.20,20,20	92.50%
18,20,20,10	93.66%
18,20,20,1	95.20%

For classification, the selected FFANN model was trained and tested for all the three types of DWT transformation methods used. Four layered FFANN model with back-propagation is used as the proposed model. The input layer has 18 neurons; two hidden layers have the number of neurons equal to number of Urdu words (classes) to be recognized, and an output layer with one neuron. The transfer function for the input and hidden layers used is Hyperbolic Tangent Sigmoid Transfer Function,and the Linear Transfer Function for the output layer. Resilient Back-propagation (Rprop) is used for training. For stopping criteria, parameters like learning rate, error tolerance rate, and number of iterations are used. Fig.11 illustrates the block diagram of the proposed network model.

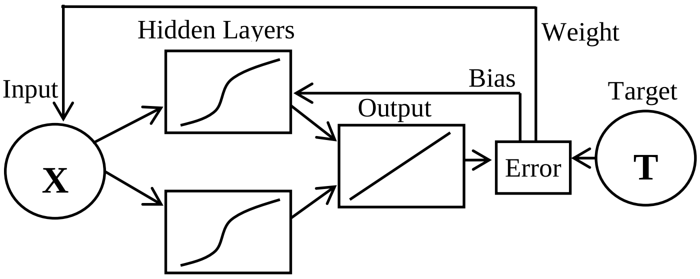


Fig.11: Block diagram of proposed FFANN model

Four layers of FFANN with back-propagation are used. The following configuration has been used for the proposed model.
Number of layers in FFANN model = 1 input layer, 2 hidden layers and 1 output layer
No. of neurons in input layer = No. of feature attributes (18)
No. of neurons in first and second hidden layer = No. of classes (Urdu words) to be classified
No. of neurons in output layer = one

Hyperbolic tangent sigmoid activation function is used in each neuron (tansig) of the input and hidden layers. Further Linear Transfer Function (purelin) is used for the output layer. Resilient Back-propagation (Rprop) has been used as training function. For the proposed system, speech samples are divided into two subsets, i.e., training subset and testing subset. For all results, division of speech samples in both the subset has been done with 50% allocation in each of the set. Random division of data function has not been adopted in this work. Further, in this work, speech samples did not add to the validation subset, as the requirementsare specific to only training and testing subsets. The following parameters of stopping criteria are taken:

Learningrate = 0.0025
Goal = 0.001
Maximum number of allowed epochs = 1000
Mean square error (MSE) functionis used as a performance measuring function for the proposed FFANN model.

MSE is a performance function that computes the average squared error between the output (*a*) of network (calculated output) and the output (*t*) of the target (desired output). Since each input is given to the NN, the output of the NN is judged with the target. The mentioned error is computed as the amount of difference between the target and the network output. The goal is to reduce the average of the sun of errors.

6.0 EXPERIMENTAL SETUP AND RESULTS

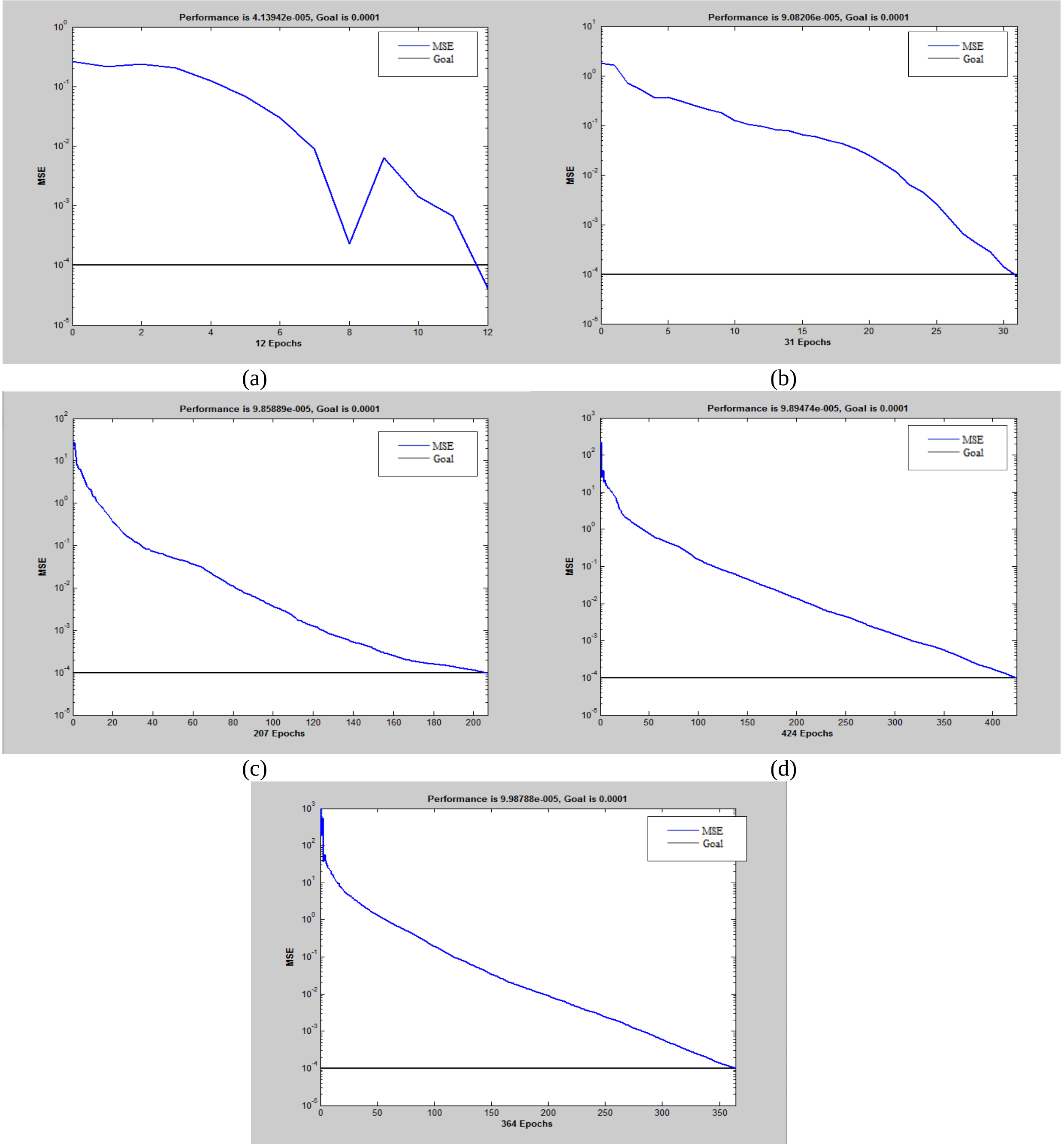
A series of experiments were conducted to find the accuracy rate of the proposed system. Inthe proposed system, speech samples of 20 Urdu words (see table 1) from tendifferent speakers were acquired. The environment effect was not considered on the proposed system. All the speech samples were recorded in a general environment. Although, the de-noising algorithm is employed, noise was not added intentionally to the speech samples. All the speech samples were recorded at different times based on the availability of speakers.

The same techniques of pre-processing and feature vector extraction were applied on all the speech samples in a single experiment. Three types of feature extraction techniques (see section 4.2) were applied. All the speech samples were divided into training and testing subset for classification. The division of samples is done such as 50% speech samples is used for training subset and the rest of 50% samples as testing subset. Experiments were performed for two, five, ten, 15, and 20 Urdu words. In each experiment, the number of sample for a single Urdu word (class) are ten(speech samples collected from tenspeakers for a single Urdu word). Therefore, on the 50% division policy, 5 samples go through training subset while the remaining fivesamples go through testing subset.After training the network, the same trained network has been used for testing the remaining 50% samples (testing subset). The accuracy is computed on the basis of predicted and target value.

Three types of experiments with fivesub-experiments were conducted. The three types of experiments are based on db-8, sym-8, and haar wavelet filters for speech recognition module to analyze the behaviour and variation in the obtained results with respect to the change in the number of classes (Urdu words), which are two, five, ten15, and 20.

6.1. Results obtained with db-8, level-5 DWT family

Experiments were performed using db-8 level 5 DWT filter on two, five, ten, 15, and 20 classes (Urdu words) with tenspeakers.



(a) (b)

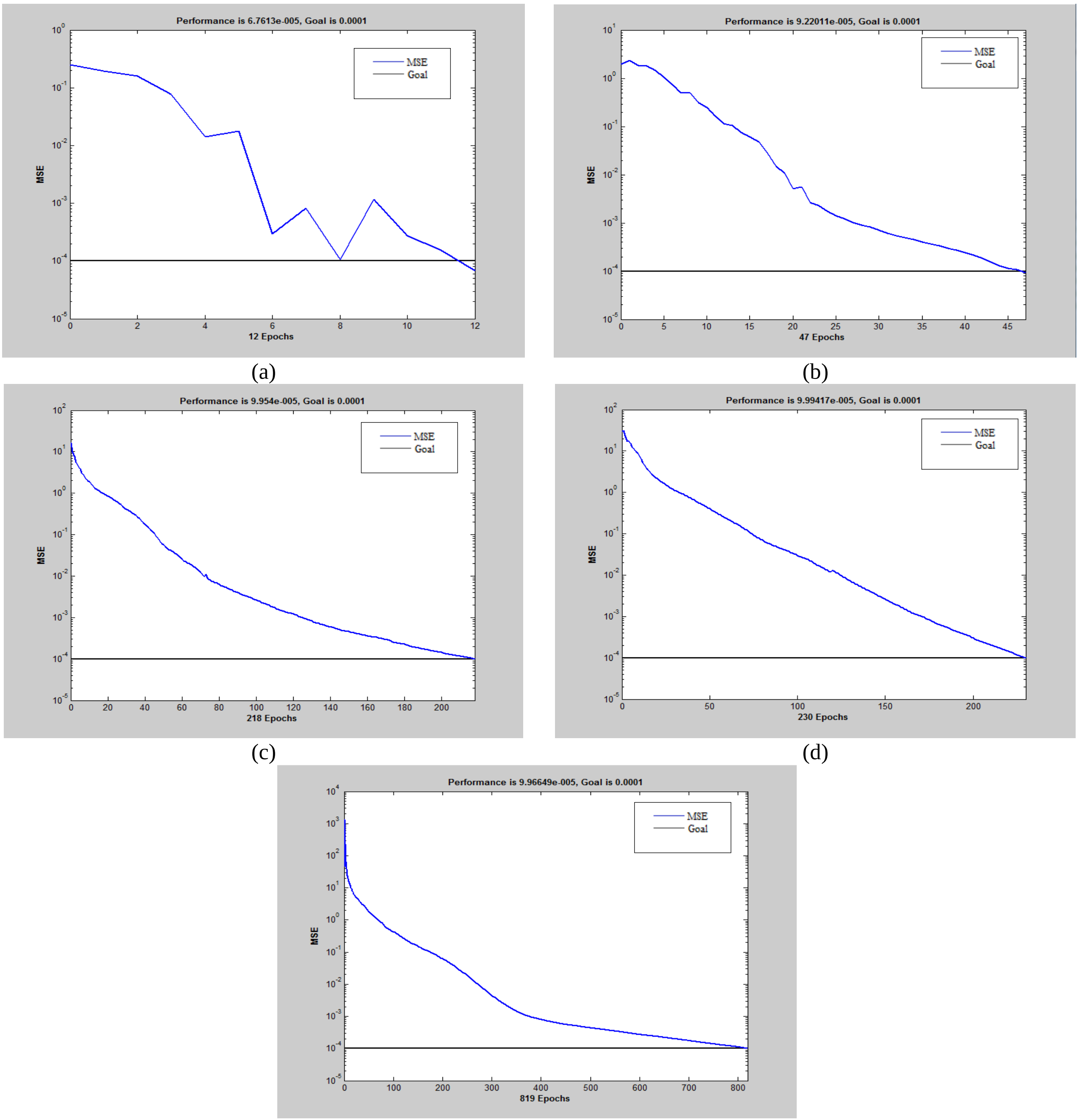
(c) (d)

(e)

Fig. 12: Net performance (a) for 2 classes using db-8 (b) for 5 classes using db-8 (c) for 10 classes using db-8 (d) for 15 classes using db-8 (e) for 20 classes using db-8

After training the network on 50% of the samples, the remaining 50% samples (reserved for testing) were used for testing. Fig. 12 (a-e) shows the network performance result for different setups of experiment using db-8 level 5 DWT family. The training process stops when MSE (blue line) reaches to goal (black line at the bottom of graphs). These graphs show the number of epochs performed during the training process and also the performance of the training process.

6.2. Results obtained with haar, level-5 DWT family



(a)

(b)

(c)

(d)

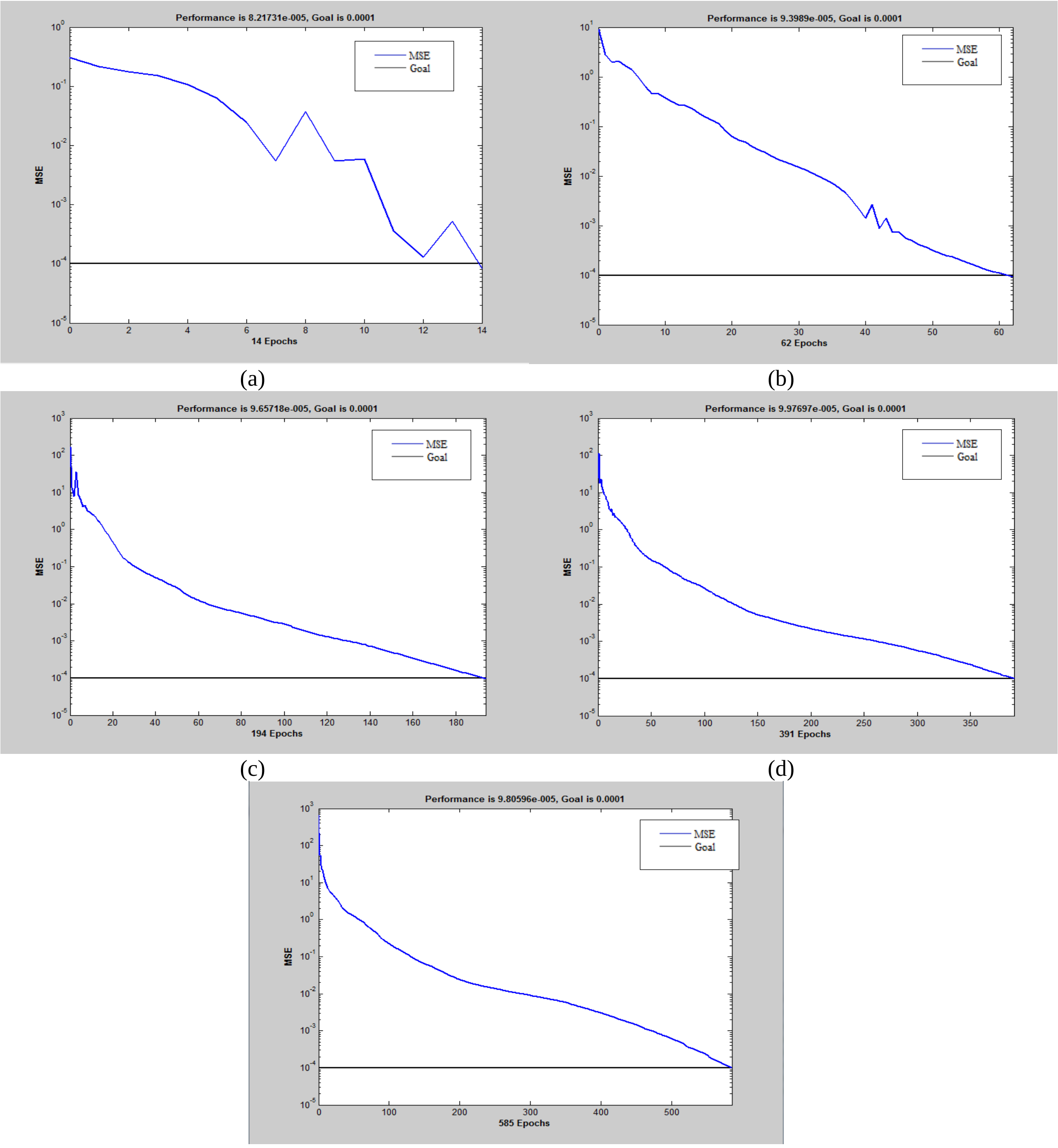
(e)

Fig. 13: Net performance (a) for 2 classes using haar (b) for 5 classes using haar (c) for 10 classes using haar (d) for 15 classes using haar (e) for 20 classes using haar

The same setup, as was used for db-8, has been used for haar level-5 DWT filter too. Fig. 13 (a-e) shows the network performance results for different setup of experiments using haar level-5 DWT family.

6.3. Results obtained with sym-8, level-5 DWT family

Experiments were also performed using sym-8 level 5 DWT filter on two, five, ten, 15, and 20 classes (Urdu words) with 10 speakers, as done previously for db-8 and haar level-5 DWT filter. The same setup has been used for sym-8 level-5 DWT filter too. Fig. 14 (a-e) shows the network performance results for different setups of experiments using sym-8 level 5 DWT family.



(a) (b)

(c) (d)

(e)

Fig. 14: Net performance (a) for 2 classes using sym-8 (b) for 5 classes using sym-8 (c) for 10 classes using sym-8 (d) for 15 classes using sym-8 (e) for 20 classes using sym-8

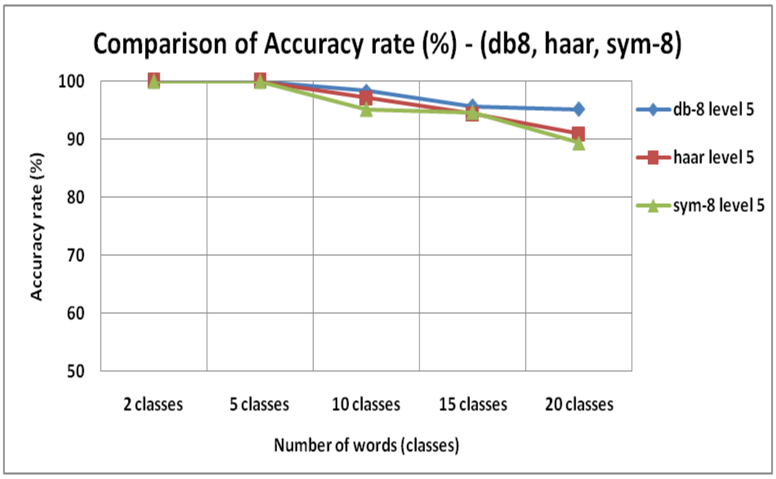


Fig. 15: Entire comparison of accuracy rate in percentage for (db-8, haar, sysm-8) lev-5 DWT filter

Table 3. Words from Pashto and Persian languages

Pashto	Persian	English equivalent	No. of samples recorded
Yaw	Yek	One	10
Dwa	Do	Two	10
Dray	Se	Three	10
Chloor	Chahar	Four	10
Pinza	Panj	Five	10
Shapag	Shesh	Six	10
Owa	Haft	Seven	10
Ata	Hasht	Eight	10
Nah	Noh	Nine	10
Las	Dah	Ten	10
Yaw-Las	Yazdah	Eleven	10
Dwa-Las	Davazdah	Twelve	10
Dyar-Las	Sizdah	Thirteen	10
Chwar-Las	Chahardah	Fourteen	10
Pinza-Las	Panzdah	Fifteen	10
Shpáarres	Shanzdah	Sixteen	10
Owa-Las	Hefdah	Seventeen	10
Ata-Las	Hejdah	Eighteen	10
Zama num	Hastam	My name	10
Shai	Drust	Right	10

6.4. Accuracy

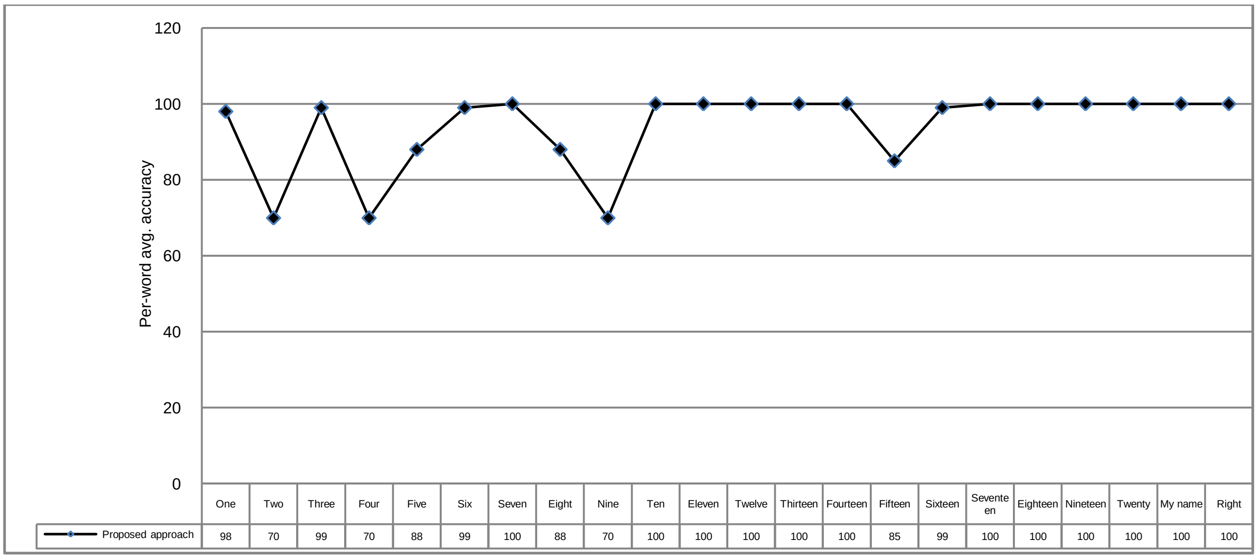
Fig. 15 combines and compares the entire system's prediction percentage accuracy rate for all the DWT filters used.From Fig.15, it is evident that for two and five classes (Urdu words), the proposed system shows 100% accuracy for all the DWT filters. Further, db-8 level-5 DWT filter shows 98.40%, 95.73%, and 95.20% accuracy rate for 10, 15, and 20 classes (Urdu words), respectively. Haar level-5 DWT filter shows 97.20%, 94.40%, and 91% accuracy for 10, 15, and 20 classes (Urdu words), respectively. Similarly sym-8 level-5 shows 95.20%, 94.67%, and 89.40% for the same number of classes (Urdu words), respectively.

From this scenario, it can be concluded that db-8 level-5 outperforms thehaar and sym-8 DWT filters for 10, 15, and 20 class problem. It also reveals that db-8 level-5 shows more consistency and accuracy as compared to others. Therefore, the best results obtained from the proposed system are using db-8 level-5. Hence, this work recommends db-8 level 5 DWT family for any speech recognition problem which may use DWT analysis. As shown in Fig.15, db-8 level-5 has the minimal amount of descent when the number of words increases. Hence db-8 level-5 performs better for a large database of speech samples.

7.0 Experiments on other oriental languages and comparison with baseline

We have conducted a detailed experiment of the proposed approach using the Urdu. This section extends the experiment presented in Section 6 by applying the proposed approach using a database from Pashto and Persian languages. We use 20 words each ofthese languages. This makes our database to constitute a total of 60 unique words. Table 3 lists the words from Pashto and Persian. The words used for the Urdu language have already been mentioned in Table 1. Further to add speaker independence and diversity to the complete dataset, these 60 words are repeatedly added to the dataset using ten different subjects (i.e., 10 male and 10 females). This makes our dataset to constitute600 samples.

To evaluate the proposed approach on the complete dataset of the oriental language, we calculate the accuracy for each of the word in the dataset. The words are repeated 4 times for recognition by 2 different subjects (one male and one female). The accuracy is computed as the average of all the 4 attempts. Fig. 16 shows the per-word average accuracy for the items in the database. Based on the results presented in Fig. 16, the average accuracy of the proposed approach for the given database of the oriental language words is 93.91%.



	One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten	Eleven	Twelve	Thirteen	Fourteen	Fifteen	Sixteen	Sevente en	Eighteen	Nineteen	Twenty	My name	Right
Proposed approach	98	70	99	70	88	99	100	88	70	100	100	100	100	100	85	99	100	100	100	100	100	100

Fig. 16: Per-word average accuracy

To compare the proposed approach with a baseline method, we have opted for the method presented in [43]. The choice of [43] is made based on its close relevance with the problem statement. The work we have proposed deals with isolated words of the oriental languages, which are the same as addressed by the approach in [43]. However, the work in [43] only addresses Bangla language. This makes [43] a baseline approach for the proposed work in this paper. We have replicated the approach in [43] for our dataset, and have shown the comparison in Fig. 17.

The approach in [43] utilized a databank of ten words recorded using ten speakers. However, the proposed work in this paper has a databank of 60 unique words each repeated ten times, making the complete dataset consisting of 600 samples. In addition to this, the databank used in this paper consists of three different languages, making the problem at hand more complex. As shown in Fig. 17, the proposed approach performs better in accuracy for all the words except for the two words, i.e., six and seven. The results in Fig. 17 are averaged over 4 iterations and consider the three orientallanguages while calculating the accuracy. Taking the complete dataset into consideration, the proposed approach has an overall accuracy of 93.91%, whereas, the accuracy of baseline method is 87.95%. The proposed approach has performed low for the words Dwa, Do, Chloor, Chahar, Nah, and Noh. It may be noticed that the baseline approach also performs lowerin recognizing these words. However, the overall accuracy of the proposed approach is better than the baseline. During the multiple iterations for recognising the isolated words, the proposed approach fails in too few of the iterations. The accuracy of the baseline approach is reported 96.332% for the Bangla language. This is due to the smaller databank. As the dataset gets larger, the chances of misclassification also increase. This is demonstrated in this work where the same approach used for Bangla language achieves low accuracy for the words from the oriental languages with a databank of 60 words and 600 samples.

The key strength of the proposed approach is the hybrid of DWT and FFANN. The DWT has the capability to capture both the frequency and location information as compared to the Fourier transform. On the other hand, the FFANN has the advantage of rapid training and the flexibility of addition/removal of training samples. This makes them an ideal combination. The isolated word recognition in this caseby using the combination of DWT and FFANN has proved to yield better results as compared to the baseline approach. The use of DWT in the proposed approach enables us to separate the fine details in a given signal. Theoretically, this is done by using small wavelets for fine details in the signal. For uneven signals, very large wavelets can be used for identification. Additionally, the DWT has an advantage of avoiding any degradation while de-noising a signal.

Most of the previous works, including the work in [43], have used Mel Frequency Cepstral Coefficient (MFCC) analysisfor feature extraction before the signal is presented to the ANN. The use of MFCC is based on an assumption that the fundamental frequency is much lower than the frequency components of the linguistic message. This is required to exclude the fundamental frequency and the harmonics. For the real-world scenarios, the aforementioned assumption is not true for the female speakers. This decreases the accuracy of the speech recognition systems utilizing MFCC.

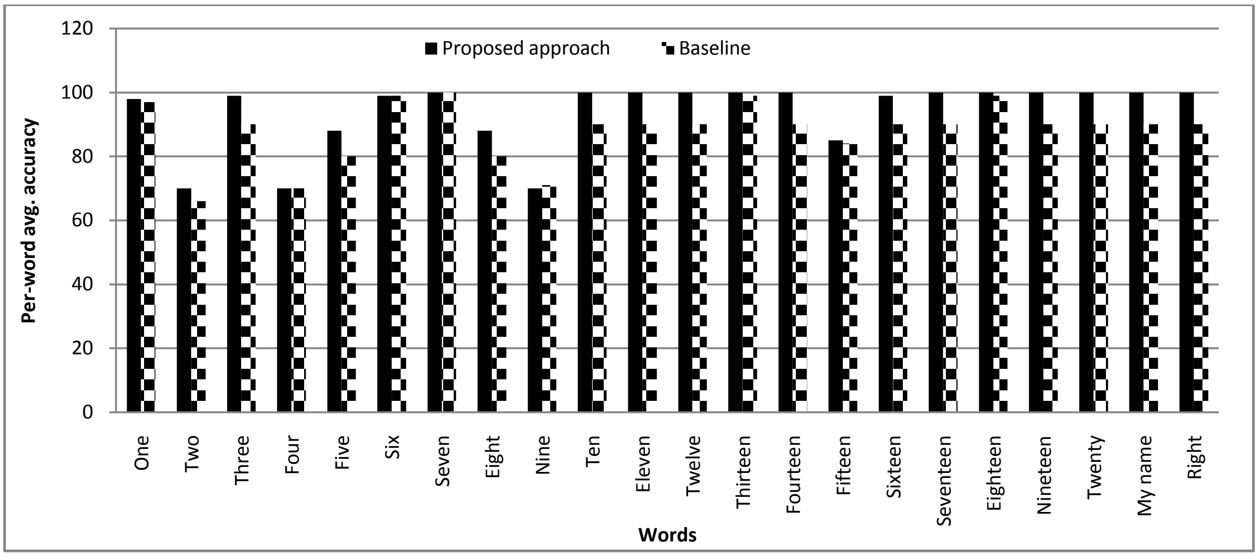


Fig. 17: Comparison with the baseline approach

8.0 DISCUSSION

The experiments section of this paper has presented a detailed work on the utilization of ANN combined with the DWT to recognize isolated words from the oriental languages. We start by pre-processing the recorded speech samples so that these can be normalized before presentation to the ANN. It was noticed that under regular conditions the initial 200 milliseconds region was classified as the unvoiced portion of speech samples. The speech samples were divided into three portions after the normalization step. The reason for this was to group a sample in three portions in a way that they have 33% region in common. The DWT was used to change the input speech signal to the frequency domain from the time domain. The major reason for using DWT for feature extraction is its capability to capture both the frequency and the location information as compared to the Fourier transform. On the other hand, the FFANN has the advantage of rapid training and flexibility of addition/removal of the training samples. For feature extraction, the three portions of the speech samples are independently decomposed using DWT. For decomposing the speech signal haar, db-8, and sym-8 are used. The decomposition is performed ofup to five levels.

Once the features are extracted from the input sample, we employed ANN as classifier. ANNs are employed because of their parallel distributed processing capability, error stability, distributed memories, and distinguishing ability for pattern learning. In the experiments, we have tested 9 variations of the ANN architectures as mentioned in Table 2. These variations are based on the number of hidden layers,the number of neurons in the hidden layer, and the number of neurons in the output layer.The experiments show better performance of the NN architecture with 18 neurons in the input layer, 20 neurons each in the 2 hidden layers, and one neuron in the output layer. For the training of ANN,Rprop was used. The system showed 100% accuracy for all the DWT filters. Further, db-8 level-5 DWT filter showed 98.40%, 95.73%, and 95.20% accuracy rate for 10, 15, and 20 classes, respectively. Haar level-5 DWT filter showed 97.20%, 94.40%, and 91% accuracy for 10, 15, and 20 classes, respectively. Similarly sym-8 level-5 showed 95.20%, 94.67%, and 89.40% for the same number of classes, respectively.

In the experiments, we have used two more oriental languages, i.e., Pashto and Persian. The accuracy for the complete dataset is calculated. The dataset consists of 600 samples with 60 unique words. The per-word accuracies of the dataset range from 70% to 100%. We see low accuracy rates for the words having similar pronunciation in Urdu, Pashto, and Persian languages. Example of such words includes: Do, Dwa, Chaar, Chloor, Chahar, and Nau, Nah, Noh, from Urdu, Pashto, and Persian languages, respectively.

9.0 CONCLUSION

This work has presented a system for speaker independent speech recognition of isolated words from the oriental languages. The proposed work combined both the DWT with FFANN to recognize isolated words. This goal was achieved with speech signal capturing, by creating database of speech samples and then by applying pre-processing techniques. For features extraction, the speech signal was decomposed using DWT with up to 5 levels, using db-8, haar, and sym-8. After decomposition, the energy on the high pass filter and low pass filters of every buffer segment was calculated and combinedwith all the 18 results in feature vectors of the speech sample.Classificationwas performed by 4 layered FFANN model with Rprop back-propagation. Experiments were performed using db-8, haar, and sym-8 level 5 DWT filter, respectively on 2, 5, 10, 15, and 20 classes with 10/20 speakers on 50% division policy. The proposed system shows100% accuracy for all the DWT filters used for twoand fiveclasses. Further, the results revealed that the db-8 level-5 outperformshaar and sym-8 DWT filters in accuracy for ten, 15, and 20 class problem.The proposed speech recognition mechanism can be used for speech recognition based on phonemes for large vocabulary and provides a communication interface to illiterate people of the subcontinent so that they can use voice commands in their native language. On the application side, this work can be extended to the isolated word recognition system for any oriental language.

REFERENCES

[1] J. Tebelskis, "Speech recognition using neural networks", PhD diss., Carnegie Mellon University, 1995.

[2] Anwar et al., "A Survey of Automatic Urdu language processing", *In proceeding of IEEE International Conference on Machine Learning and Cybernetics*, China, 2006 , pp. 4489-4494.

[3] S. Hussain, “Letter-to-sound conversion for Urdu text-to-speech system”,*In proceeding of Association for Computational Linguistics Workshop on Computational Approaches to Arabic Script-based Languages*,Switzerland, 2004,pp.74-79.

[4] Moohebat, M., Raj, R.G. , Kareem, S.B.A., Thorleuchter, D., “Identifying ISI-indexed articles by their lexical usage: A text analysis approach”, *Journal of the Association for Information Science and Technology*, Vol. 66, No. 3, pp. 501–511. doi: 10.1002/asi.23194.

[5] I. Daubechies, Ten lectures on wavelets. Vol. 61, Philadelphia, PA: Society for industrial and applied mathematics, 1992.

[6] Mallat et al. ,"A theory for multiresolution signal decomposition: the wavelet representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, 1989, pp. 674-693.

[7] Y. Meyer, Ondelettesetopérateurs: Ondelettes, Vol. 1, 1990, Hermann.

[8] A. A. Raza et al. ,"Design and development of phonetically rich Urdu speech corpus",*In proceeding of IEEE International Conference on Speech Database and Assessments*, Taiwan, 2009, pp. 38-43.

[9] A. W. Abbas et al. ,"Pashto Spoken Digits database for the automatic speech recognition research",*In proceeding of 18th IEEE International Conference on Automation and Computing*,United Kingdom, 2012,pp. 1-5.

[10] F. Naz et al. ,"Urdu Part of Speech Tagging Using Transformation Based Error Driven Learning", World Applied Sciences Journal, vol. 16, no. 3, 2012, pp. 437-448.

[11] Yeow, W.L., Mahmud, R., Raj, R.G., "An application of case-based reasoning with machine learning for forensic autopsy", *Expert Systems with Applications*, Vol 41, No. 7, 2014, pp. 3497-3505, ISSN 0957-4174, http://dx.doi.org/10.1016/j.eswa.2013.10.054. (http://www.sciencedirect.com/science/article/pii/S0957 417413008713).

[12] J. D. Wu et al. ,"An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network", Expert systems with applications, vol. 36, no. 3, 2009, pp. 4278-4286.

[13] S. K. Hasnain,"Confusion Matrix Validation of an Urdu Language Speech Number System",*In proceeding of the International Conference on Artificial Intelligence*, USA,2008, pp. 67-73.

[14] E. S. Hasnain et al. ,"Cohesive Modeling, Analysis and Simulation for Spoken Urdu Language Numbers with Fourier Descriptors and Neural Networks", Ubiquitous Computing and Communication Journal, vol. 3, no. 2, 2007,pp. 196-211.

[15] S. K. Hasnain et al. ,"Recognizing spoken Urdu numbers using fourier descriptor and neural networks with Matlab"*In proceeding of IEEE Second International Conference on Electrical Engineering*, Pakistan,2008, pp. 1-6.

[16] S. K. Hasnain et al. ,"A Speech Recognition System for Urdu Language",*In proceeding of IEEE Multi-Topic Conference*, Pakistan, 2008,pp. 74-78.

[17] Y. A. Alotaibi, "Investigating spoken Arabic digits in speech recognition setting. Information sciences", vol. 173, no. 1, 2005, pp. 115-139.

[18] G. E. Dahl et al. ,"Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition", IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 1, 2012, pp. 30-42.

[19] B. Baldi, Student Solution Manual for The Practice of Statistics in the Life Sciences. 2011, WH Freeman.

[20] R. D.Maesschalck et al. ,"The mahalanobis distance",Chemometrics and Intelligent Laboratory Systems, vol. 50, no. 1, 2000,pp.1-18.

[21] A. Bultheel, Wavelets with applications in signal and image processing. Ed. 2, 2003, SPIE.

[22] A. Ukil, Intelligent systems and signal processing in power engineering. 2007, Springer Verlag.

[23] H. Y. Chan et al. ,"Discriminative pronunciation learning for speech recognition for resource scarce languages",*In Proceedings of the 2nd ACM Symposium on Computing for Development*, USA,2012, pp. 12.

[24] Z. Halim et al., "A Kinect-Based Sign Language Hand Gesture Recognition System for Hearing-and Speech-Impaired: A Pilot Study of Pakistani Sign Language", Assistive Technology, vol. 27, no.1, 2015, pp. 34-43.

[25] J. Sherwani et al., "Speech vs. touch-tone: Telephony interfaces for information access by low literate users," IEEE International Conference on Information and Communication Technologies and Development, 2009, pp. 447-457.

[26] A. Mumtaz et al., "Literacy Efforts In Pakistan: Need For Altering The Course Of Action," The AYER, vol. 3, 2015, pp.300-307.

[27] A. Blandford, "Google, Public Libraries, and the Deep Web," Dalhousie Journal of Interdisciplinary Management, vol. 11, 2015.

[28] D. Zeng, "AI Ethics: Science Fiction Meets Technological Reality," IEEE Intelligent Systems, vol. 30, no. 3, 2015, pp. 2-5.

[29] A. Gandomi at al.,"Beyond the hype: Big data concepts, methods, and analytics," International Journal of Information Management, vol 35, no. 2, 2015, pp. 137-144.

[30] B. Rehman et al.,"ASCII Based GUI System for Arabic Scripted Languages: A Case of Urdu," International Arab Journal of Information Technology, vol. 11, no. 4, 2014.

[31] M. Varouqa et al.,"WIT: Weka interface translator," International Journal of Speech Technology, 2015, pp.1-13.

[32] R. Kumar et al.,"Fuzzy-Membership Based Writer Identification from Handwritten Devnagari Script," ournal of Information Processing Systems, 2014.

[33] D. Sakata et al.,"Education System to Learn the Skills of Management Decision-Making by Using Business Simulator with Speech Recognition Technology," Industrial Engineering & Management Systems, vol. 13, no. 3, 2014, pp. 267-277.

[34] A. Graves et al., "Towards end-to-end speech recognition with recurrent neural networks," Proceedings of the 31st International Conference on Machine Learning, 2014.

[35] O. Abdel-Hamid et al., "Convolutional neural networks for speech recognition," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no.10, 2014, pp. 1533-1545.

[36] E. Arisoy et al., "Converting neural network language models into back-off language models for efficient decoding in automatic speech recognition," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no.1, 2014, pp. 184-192.

[37] Raj, R.G., Abdul-Kareem, S., “Information Dissemination And Storage For Tele-Text Based Conversational Systems' Learning”, *Malaysian Journal of Computer Science*, Vol. 22(2):2009. Pp. 138-159.

[38] G. Tamulevičius et al., "Hardware accelerated FPGA implementation of Lithuanian isolated word recognition system," ElektronikairElektrotechnika, vol. 99, no. 3, 2015, pp. 57-62.

[39] L. Linh et al., "MFCC-DTW Algorithm for Speech Recognition in an Intelligent Wheelchair," FMBE Proceedings, vol. 46, 2015, pp. 417-421.

[40] H. Ali et al., "Linear discriminant analysis based approach for automatic speech recognition of Urdu isolated words," Communication Technologies, Information Security and Sustainable Development, 2014, pp. 24-34.

[41] H. Ali et al., "DWT features performance analysis for automatic speech recognition of Urdu," SpringerPlus, vol. 3, no. 1, 2014, pp. 1-10.

[42] J. Ashraf et al., "Speaker independent Urdu speech recognition using HMM," IEEE 7th International Conference on Informatics and Systems (INFOS), 2010.

[43] M. A. Hossain, M. M. Rahman, U. K. Prodhan and M. F. Khan, “Implementation Of Back-Propagation Neural Network For Isolated Bangla Speech Recognition,” International Journal of Information Sciences and Techniques , Vol. 3, No.4, pp. 1-9, 2013.

[44] Z. Halim, A. R. Baig, K. Zafar, "Evolutionary Search in the Space of Rules for Creation of New Two-Player Board Games," International Journal on Artificial Intelligence Tools, Vol. 23, No. 2, pp. 1-26, 2014.