Vehicles Sound Recognition Using Wavelet Transform, Fast Fourier Transform and Neural networks

Dr. Yassir Mohammed HajHamad
Karary university,Collge of Computer science&Information Technology, Sudan

Dr. Iman Abdrahman A.Mally
Khartoum university , College of Engineering ,Sudan

Abstract:
Vehicles may be recognized from the sound they make when engine moving from their acoustic signature. In this paper, investigated two feature extraction methods for acoustic signals, Fast Fourier Transform(FFT) and Wavelet Packet Transform(WPT) and used for recognition. To reduce the amount of features extracted used Principal Component Analysis algorithm. And used Artificial neural networks as a classifier for Vehicles types.
Keywords: *Feature Extraction, FFT, WPT, PCA, ANN.*

المستخلص :
يمكن التعرف على المركبات من خلال بصمه صوت المحرك ومن خلال هذه الورقه تم تشخيص اشارات بصمة الصوت باستخدام اثنين من طرق استخلاص معالم الصوت وهما تحويل فورير السريع وتحويل حزم المويجات في عمليه التعرف ولتقليل حجم استخلاص المعالم تم استخدام خوارزميه تحليل المتجهات وتم استخدام الشبكات العصبونيه الاصطناعيه للتعرف على انواع المركبات.
الكلمات المفتاحية: استخلاص المعالم ، تحويل فورير السريع , تحويل حزم المويجات , خوارزميه تحليل المتجهات , الشبكات العصبونيه الاصطناعيه.

1.Introduction:

Pattern Recognition theory can be a good approach to the sound classification problem. In pattern recognition, objects are identified on the bases of some attributes or features(1). The selection of a set of features that is capable of distinguishing between classes are the most critical step in audio classification system design (2). All vehicles emit characteristic sounds when engine moving. These sounds may come from various sources including rotational parts, vibration in the engine, wind effect, gears, friction between the tires and fans. Similar vehicles working in comparable conditions would have a similar acoustic signature that could be used for recognition. Acoustic signature plays a crucial role in the military operations and several systems have been proposed for vehicle recognition. Classification of vehicles based on acoustic signals can be employed effectively in battlefield surveillance, traffic control, and many other applications. The classification performance depends on the selection of signal features that determine the separation of different signal classes. Rest of the paper is organized as follows: section 2 describes recent methods to extract features from the vehicle acoustic signals. Section3presents efficient and useful technique for classification phase. ection4concludes the paper with a summary and comparison among the different methods of feature extraction and classification.

2. Feature Extraction Techniques

Extraction of characteristic features(acoustic signature) is the most significant and crucial task in vehicle recognition. For classification of a vehicle, a set of characteristic features from sound generated by vehicle is extracted to label that vehicle to one of the predefined classes. Acoustic signals have very few representations and we must find the most important coefficients or characteristic, which contain information that will be used to discriminate among input classes. This set of characteristic features is known as a feature vector(3). Feature extraction followed by feature reduction technique, this technique used to extract the acoustic features of the vehicles and to uncover the most important acoustic features of an unknown signal that differentiate it from others. A popular statistical tool for dimensionality reduction technique is Principal Component Analysis(PCA). The objective of PCA is to find principal eigenvectors of the covariance matrix of the set of signals. These eigenvectors can be considered as characteristic feature vector which is used to characterize the variation between signals(4). Three main domains in which features can be generated are time domain, frequency domain and time- frequency domain. The feature extraction in time domain is simplest method for feature extraction because no transformation is required so computationally it become less complex(energy envelop, zero crossings).

2.1 Feature Extraction in Frequency Domain

The feature extraction in frequency domain, acoustic signal waveforms generated by vehicle appears as narrow band harmonic components. The information provided by these components, is used to construct the characteristic features for a particular vehicle. Characteristic features are generated by using low frequency components present in the acoustic signal because most of the sound produced by the vehicles is due to their engine rotating parts which rotate and reciprocate in a low frequency and they are strongest and suffer least attenuation(5). Feature generation methods based on frequency domain such as Fast Fourier Transform(FFT) and Power spectral density(PSD) are commonly used in vehicle recognition and classification(6).

2.2Preprocessing Operations

Let Y be the sound recorded to construct feature vector. Y is divided into multiple sound frames Yi , where Yi is a basic unit of analysis. Preprocessing is the basic step in which DC bias which may be caused by the device during sampling, is removed by applying zero mean to the Yi.

$$\mathrm{Yi(n)} = \mathrm{yi(n)} - \frac{1}{N} * \sum_{n=1}^{N} \mathrm{yi(n)} \qquad .(1)$$

Then we get:

$$\mathrm{yij} = \mathrm{yij(old)} - (1/\mathrm{n})^{*} \sum_{\mathrm{k=0}} \mathrm{yik} \qquad .(2)$$

Where K=1,2,……..,n.

Now this zero mean sound is converted into frequency domain by applying FFT. the environmental noise removed by using Finite Impulse Response(FIR) filter:

$$y[n] = \sum^{n} h[k].\,y[n-k]. \qquad .(3)$$

Where h[k] is an Impulse Response, where k=1,2,…,n.

y[n]: the input frame of Yi , Y[n]: the filter signal of frame Yi

in the recording process, we recorded acoustic signals or the quasi-periodic signals for each class in 120 (seconds), to increase the dataset of sound waves, we divided each length of signal into 12frames with 10 (seconds), for each one. Figure (1) illustrates the car sound waves in the first case (in time domain and frequency domain), before divided into frames.

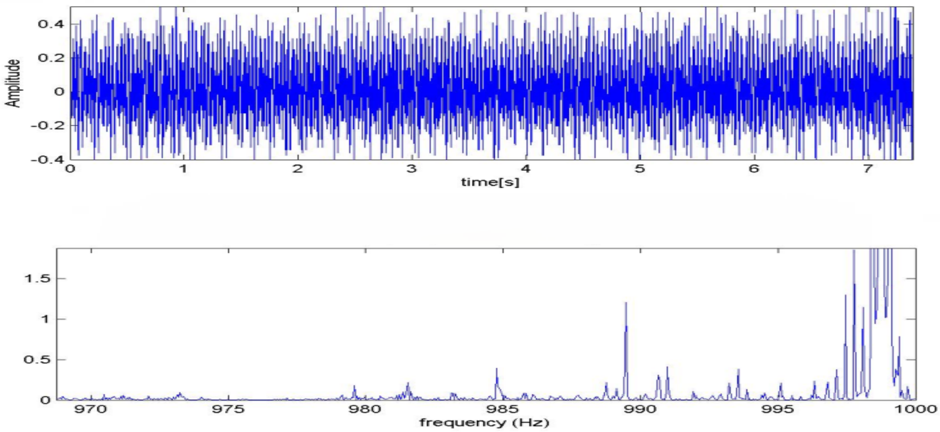


Fig (1): The car sound wave in Time Domain and Frequency Domain

The signal length divided into 12 Frames, the length of each one was 10 (seconds), so the dataset training becomes 120 samples for class1 (car), and so the others classes. Fig (2) illustrates Frames after framing process. According to the sampling rate, the size *L* of a proper window should be selected firstly, such as 256, 512, 1024, 2048, and 4096 sampling points. In this research we used

512 sampling points, in order to obtain N short-time series {x1, x2, . . . , xN}.

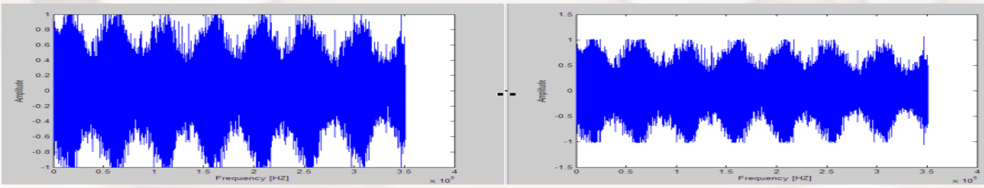


Fig (2): The Helicopter sound wave before (FIR).The Helicopter sound wave After (FIR).

2.3 Feature extraction using FFT

Magnitude of frequency spectrum used to generate feature vector by FFT for a window size 512, without overlapping can be written as:

$$\text{Each vehicle has a unique} = \text{FFT}(yi) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(4)$$

3.Feature Extraction in Time-Frequency Domain

The techniques used to extract features in time-frequency domain are Short Time Fourier Transform(STFT) and Wavelet Transform(WT)(7). Wavelet Transform provides multi-resolution time-frequency analysis. It is the projection of a signal onto the wavelet. Wavelet is a series of functions Ψab(t)derived from a base function ψ(t) by translation and dilation[8].

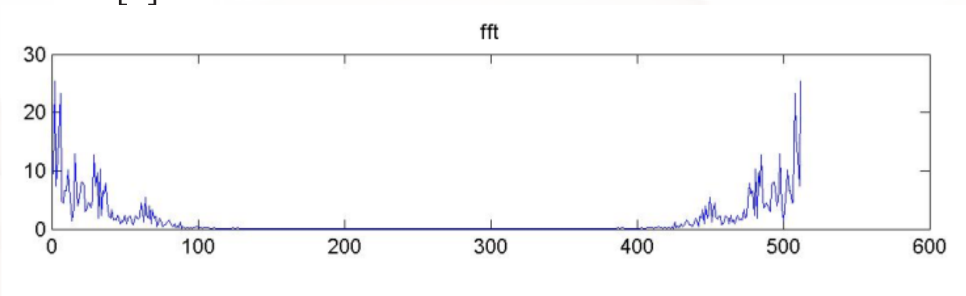


Fig (3): car with spectrums using FFT.

4.Wavelet Packet Transform:

This transform is computationally efficient. The main concept used in this technique is to extract the set of features from the class of signals emitted by a certain vehicle. The set of features is obtained by calculating the inherent energies in the blocks of the wavelet packet coefficients of the signal, each of which is related to a certain frequency band. Wavelet packet transform of a signal yields different partitions of the frequency domain. Due to the presence of time variance property in the multi scale wavelet packet decomposition, whole blocks of wavelet packet coefficients are used instead of individual coefficients and waveforms(11). To generate feature vector, a set of signals are sliced into fragments of length L=10s and each fragment is subjected to the wavelet packet transform. Wavelet packets transform works as a bridge between the time domain and frequency domain representation of a signals[12].Wavelet packet transform can be viewed as a tree structure. The root of the tree is the time series of the vehicle sound. The next level is the result of one step of wavelet transform. Subsequent levels(3 levels) in the tree are obtained by applying the wavelet transform to the low and high pass filter results of the previous step's wavelet transform. The Branches of the tree are the blocks of coefficients. Each block represents a band of frequency. Feature extraction of acoustic signals is based on the energy distribution of the block coefficients of wavelet packet transform. The wavelet packet transform is applied for acoustic signals then the energy of each block coefficients of the three levels of db4 is calculated

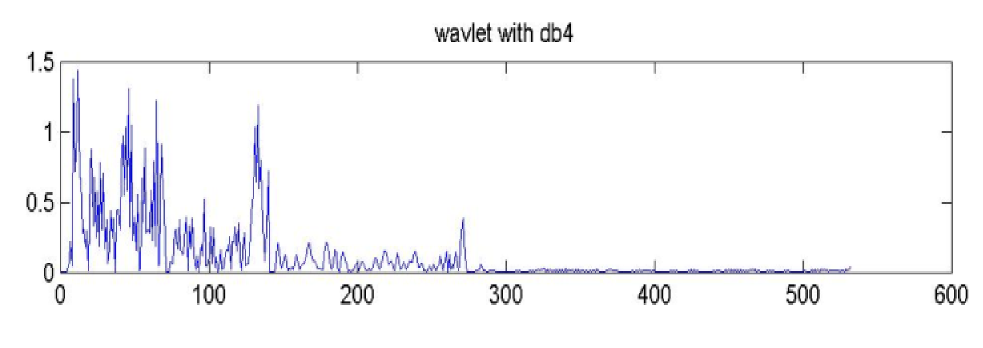


Fig (4): Spectral Analysis Using Wavelet for Helicopter.

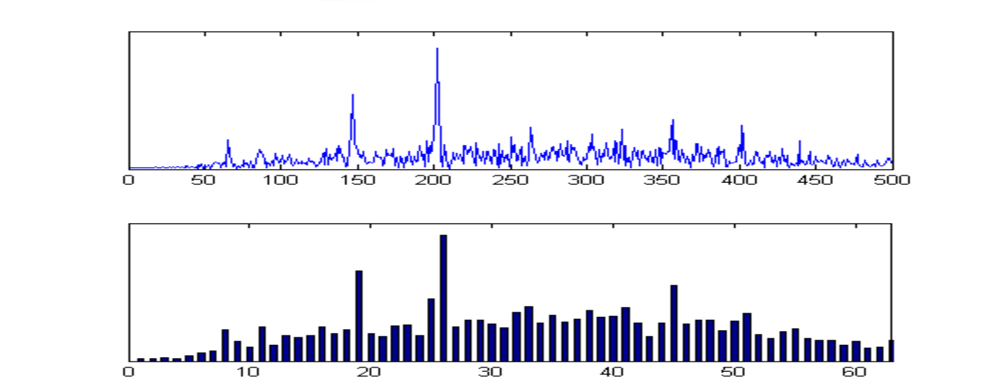


Fig (5): Top: Fourier spectrum of the car signal .Bottom: Energies in the blocks of Wavelet packet coefficients of the fourth order in three level of the wavelet packet transform.

5.Feature Reduction (Selection).

Principal component analysis is a quantitatively rigorous method for achieving this simplification. The method generates a new set of variables, called principal components. Each principal component is a linear combination of the original variables. All the principal components are orthogonal to each other, so there is no redundant information. The principal components as a whole form an orthogonal basis for the space of the data.

6.Artificial Neural Network:

Neural Networks or artificial neural networks to be more precise represent a technology that is rooted in many disciplines, neurosciences, mathematics, statistics, physics, computer science and engineering. Neural networks find applications in such diverse fields as modeling, time series analysis, pattern recognition, signal processing and control by virtue of an important property: the ability to learn from input data(16). In general, we may identify two fundamentally different classes of network architectures (17). The neurons are organized in the form of layers. In the simplest form of a layered network, we have an input layer of source nodes that projects onto an output layer of neurons (computation nodes), but not vice versa. This network is strictly a feed forward or acyclic type. (see figure1), for the case of four nodes in both the input and output layers. Such a networks called a single-layer network. In this network, each element of the input vector p is connected to each neuron input through the weight matrix W. The neuron has a summer that gathers its weighted inputs and bias to form its own scalar output n(i). The various n (i) taken together form an S-element net input vector n. Finally, the neuron layer outputs form a column vector a. The second class of a feed forward neural network distinguishes itself by the presence of one or more hidden layers, whose computation nodes are correspondingly called hidden neurons or hidden units. The source in the input layer of the network supply respective elements of the activation pattern (input vector), which constitute the input signals applied to the neurons (computation nodes) in the second layer (i.e., the first hidden layer). A prescribed set of well-defined rules for the solution of a learning problem is called a learning algorithm. As one would expect, there is no unique learning algorithm for the design of neural networks. Basically, learning algorithms differ from each other in the way in which the adjustment to a synaptic weight of a neuron is formulated. Learning paradigm is a model of the environment in which the neural network operates. There are three major learning paradigms, each corresponding to a particular abstract learning task. These are [18]:

- Supervised learning: the weights and biases are modified in response to network targets.
- Unsupervised learning: the weights and biases are modified in response to network inputs only.
- Reinforcement learning: Data is usually not given, but generated by an agent's interaction with the environment

The Back-propagation network is an algorithm that can propagate and back propagate the network from one layer to other layers. The network consists of an input layer of source neurons, at least one middle or hidden layer of computational neurons, and an output layer of computational neurons. The input signals are propagated in a forward direction on a layer by layer basis and backward direction on a layer-by layer basis.

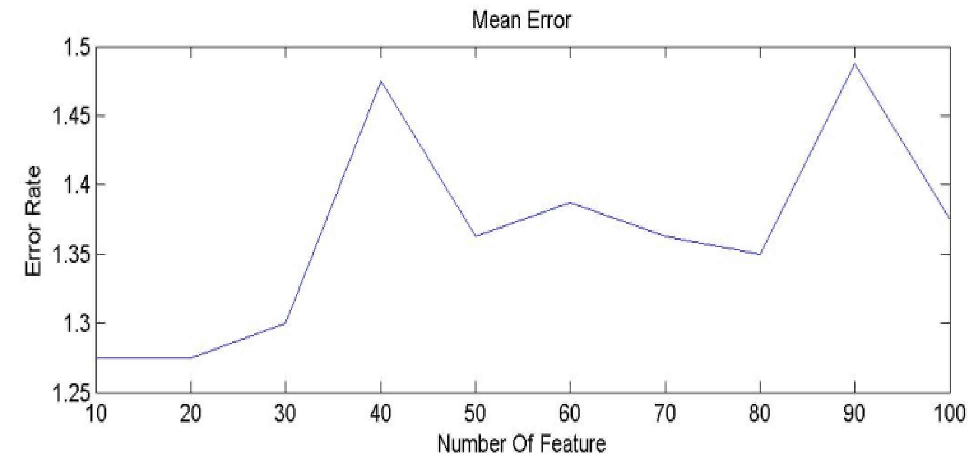


Fig (6) : The PCA Error Rates with Two Hidden Layers.

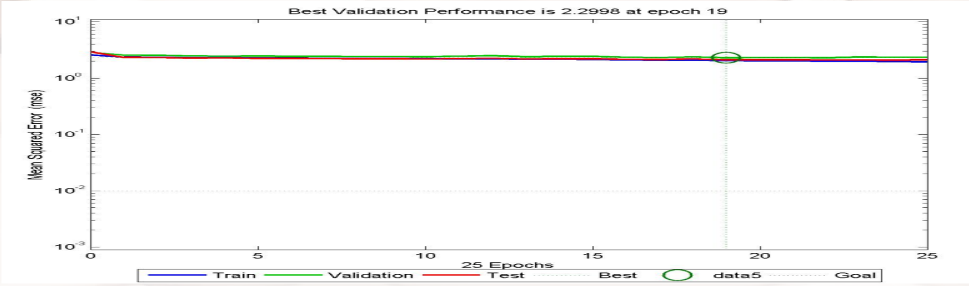
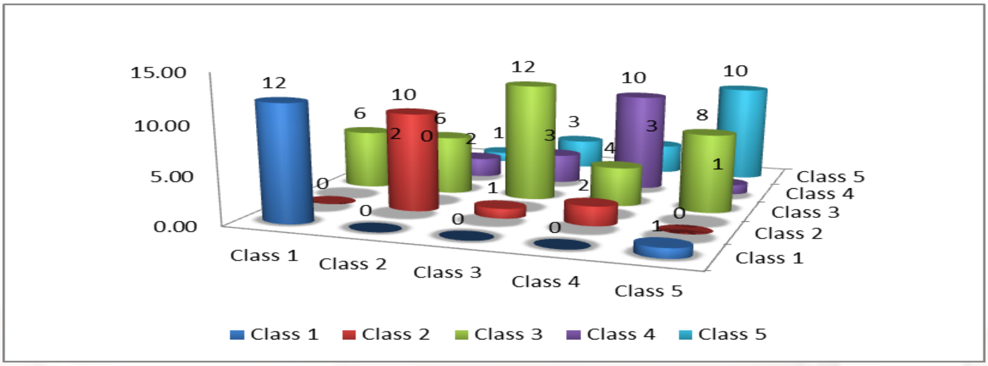


Fig (7): the PCA input20 features with one hidden layer.

Table (1): Confusion matrix with one hidden layer,PCA of 10 and FFT

	Class 1	Class 2	Class 3	Class 4	Class 5	Not classified
Class 1	12	0	6	2	0	0
Class 2	0	10	6	2	1	1
Class 3	0	1	12	3	3	1
Class 4	0	2	4	10	3	1
Class 5	1	0	8	1	10	0

Table (2): Recognition Rate With One Hidden Layer,PCA of 10 and FFT.



Conclusion

Classification of vehicles based on acoustic signals is being employed effectively in battlefield surveillance, traffic control, and many other applications, but research in this area still requires more investigation to improve the classification accuracy. Sound recognition that identifies vehicles according to their acoustic signatures was successfully designed and implemented. Sounds of five vehicles were used for training and testing the system. The data set contained 20 sound recordings for each vehicle sampled at 4960 Hz, with 8 bit resolution. An FIR noise reduction filter was used to improve the quality of input signal. For features extraction Radix -2 Fast Fourier Transform (FFT) with N=512 and Wavelet Packet Transform (WPT) with 31-sub-bands were tested. The Principal Components Analysis (PCA) was utilized as a features dimensionality reduction method. Artificial Neural Network (ANN) was employed as a classifier. Feed-forward Multi-Layers Perceptron (MLP) trained by Back-Propagation Algorithm (BPA) was used to train and classify the feature vectors of vehicle sound. Cross validation was implemented to prevent the over fitting problem and to increase the training data samples. performance was tested with different feature extraction methods, and different feature lengths of the PCA, along with different number of hidden layers of the ANN. The results showed that the highest classification accuracy gained is 94% when WPT is used as a feature extraction method with number of features reduced to 90 using the PCA and with two hidden layer ANN as illustrated in tables 5,6. This is compared to 88% classification accuracy gained when the Fast Fourier transform is used as a feature extraction method with number of features reduced to 20 and with one hidden layer Neural Network. Compared to previous studies in the field of vehicle sound recognition, the highest calcification accuracy achieved in previous studies was 90%. as illustrated in tables 3,4

Table (3): Confusion matrix with one hidden layer,PCA 20 and FFT.

	Class 1	Class 2	Class 3	Class 4	Class 5	Not classified
Class 1	19	0	0	1	0	0
Class 2	0	20	0	0	0	0
Class 3	0	0	18	0	1	1
Class 4	0	0		20	0	0
Class 5	1	0	0	1	18	0

Table (4): Recognition Rate With One Hidden Layer,PCA20 and FFT.

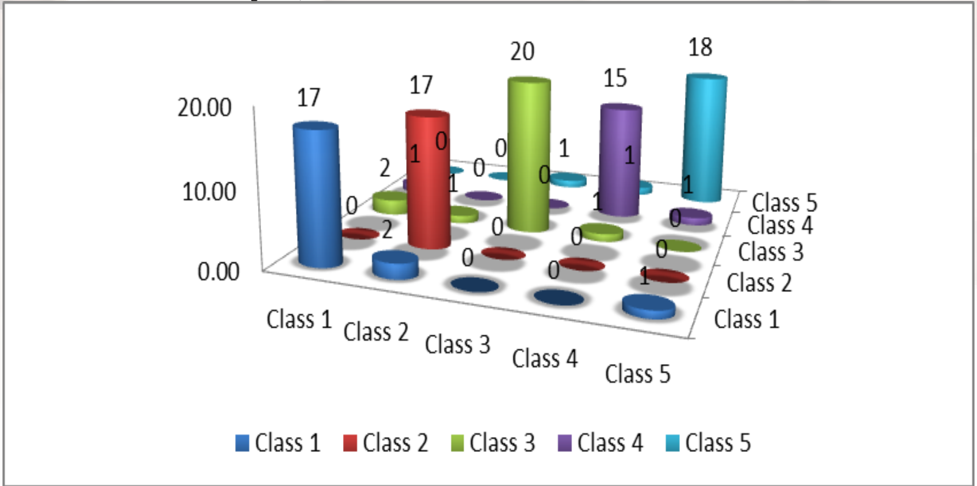
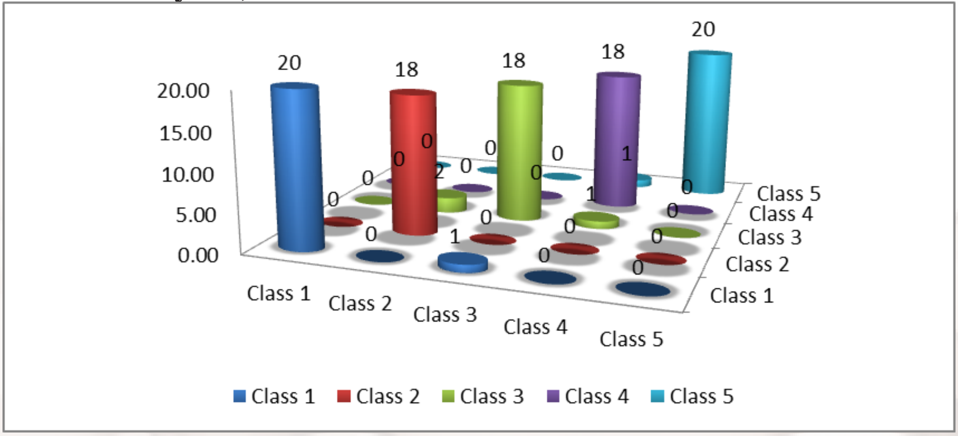


Table (5): Recognition Rate with Two hidden layer,PCA90 and WPT.

	Class 1	Class 2	Class 3	Class 4	Class 5	Not classified
Class 1	20	0	0	0	0	0
Class 2	0	18	2	0	0	0
Class 3	1	0	18	0	0	1
Class 4	0	0	1	18	1	0
Class 5	0	0	0	0	20	0

Table (6): Recognition Rate With Two Hidden Layers, PCA90 and WPT.



References

(1) K. Fukunaga, " Introduction to Statistical Pattern Recognition", San Diego, California: Academic Press, 2017

(2) [2] V. Peltonen, J. Tuomi, A. Klapuri, and Jyri, "Computational auditoryrecognition", IEEE conference on Acoustics, Speech, and Signal Processing, 2020.

(3) Liu. And Yufeng, " Comparative study of feature extraction and Classification methods for military vehicle recognition", Institute of Technology-china, 2007.

(4) H. Wu, M. Siegel, and P. Khosla, "Vehicle sound signature recognition by frequency vector principal component analysis," IEEE Trans.Instrum. Meas., vol. 48, no. 5, pp. 1005–1009, Oct. 1999.

(5) D. Lake, " Tracking fundamental frequency for synchronous mechanical diagnostic signal processing." In Proceedings of 9th IEEE Signal Processing Workshop on Statistical Signal and Array Processing, 2014

(6) R. Battiti, "Using mutual information for selecting features in supervised neural net learning." IEEE Transactions on Neural Networks, 5(4):537–550, July 1994.

(7) Y. Sun and H. Qi., "Dynamic target classification in wireless sensor networks." In Pattern Recognition, ICPR ,19th International Conference.

(8) E.M. Munich, "Bayesian subspace methods for acoustic signature recognition of vehicles." in Proceedings of the European Signal Processing Conference (EUSIPCO-04), Vienna, Austria, Sept. 2004.

(9) H. Wang, W. Yang, W.-d. Zhang, and Y. Jun, "Feature extraction of acoustic signal based on wavelet analysis." In ICESSSYMPOSIA '08: Proceedings of the 2018 International Conference on Embedded Software and Systems Symposia. Washington, DC, USA: IEEE Computer Society, 2018.

(10) A. H. Khandoker, D. T. H. Lai, R. K. Begg, and M. Palaniswami, "Wavelet-based feature extraction for support vector

machines for screening balance impairments in the elderly," vol. 15, 2007.

(11) Amir Averbuch, Valery A. Zheludev, Neta Rabin, Alon Schclar, "Wavelet-based acoustic detection of moving vehicles." Multidim Syst Sign Process, Springer Science+Business Media LLC 2017

(12) Candes, E., Romberg, J., Tao.T., "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information." IEEE Transactions on Information Theory, 2016

(13) Donoho, D., & Tsaig, Y., " Extensions of compressed sensing. " Signal Processing, 86(3), 533–548 (2016).

(14) Averbuch,A. Z., Hulata, E., Zheludev,V. A., Kozlov, "A wavelet packet algorithm for classification and detection of moving vehicles." Multidimensional Systems and Signal Processing, 12(1), 9–31. I. (2001a).

(15) Ahmad Aljaafreh, Ala Al-Fuqaha, "Multi-Target Classification Using Acoustic Signatures in Wireless Sensor Networks: A survey." Signal Processing-An International Journal (SPIJ), Volume (4): Issue (4).

(16) Howard Demuth and Mark Beale, " Neural networks Toolbox for Use with Mat lab", by the Math works, Inc, 2017.

(17) Jack M. Zurada , "Introduction to Artificial Neural Systems", 2015.

(18) Howard Demuth and Mark Beale, "Neural networks Toolbox for Use with MATLAB", by the Math works, Inc, 2018

(19) Siti Mariyam, "Soft Computing and Neural network", Produced by University of Technology Malysia, 2019.

.