iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50602

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Interactive System for Gender Classification

Mr. G. Tirupati[1], K. V. S.Sirisha[2], M. Vijaya[3], K. N. Swetha[4], G. Aruna[5]
Information Technology, GVPCEW, Visakhapatnam, India

Abstract: ***A study on the detection of gender, based on voice using artificial neural networks. In today's fast-moving world Gender classification through voice plays an important role to enhance performance in Speech recognition systems, Forensic investigation, and Marketing. The dataset has 3,168 recorded voice samples of male and female voices. The samples are produced by using acoustic analysis. The primary goal of the proposed model is to automate the system to identify gender based on the audio signals and to test the voice of a human on the spot. Multilayer Perceptron (MLP) with ReLU activation function as a model has been trained to predict gender. Nadam optimizer is used for the optimization of neural networks, K-Nearest Neighbor and Support Vector Machines are trained on the dataset of 3,168 records. The obtained best accuracy is 97% on the given dataset by MLP algorithm. An Interactive web page has been built to test the voice without interruption and to predict its gender. 125 real time samples are tested out of which model could classify every record into male/female. 19 male records are incorrectly classified.***

Keywords: Artificial Neural Networks, Multilayer Perceptron (MLP), ReLU, Nadam optimizer, acoustic analysis, K-Nearest Neighbor, Support Vector Machines.

I. INTRODUCTION

Acoustic analysis is a method of analyzing acoustic signals.

The frequency range of human voice can be used for gender classification. In general, male voices tend to have a lower frequency range than female voices. The average fundamental frequency of adult male speech is around 85 to 180 Hz, while the average fundamental frequency of female speech is around 165 to 255 Hz. Therefore, in gender classification tasks, a common approach is to extract acoustic features related to fundamental frequency such as mean, standard deviation, and more.

warbleR r package with specan function is designed for getting parameters from acoustic analysis to identify the gender of speaker. Each voice sample format is a .WAV file. The .WAV format files have been preprocesses using specan function in WarbleR r package which measures 22 acoustic parameters on acoustic signals for which the start and end times are provided. The preprocessed .WAV files are saved into a .CSV file for further training of the model.

The dataset consists of 3,168 rows and 21 columns which includes label classification of speaker as male or female.

An Interactive web page has been built using flask framework to test the gender of the speaker on the spot.

Web Pages are developed and designed using Html5, CSS, and JavaScript modules and are integrated with the system which is based on the principles of multilayer perceptron networks.

Section 2 describes the literature survey and section 4 contains the methodology of the proposed system along with model architecture and details of the framework for frontend. It also contains implementation details of the proposed system and Section 5 contains execution and results followed by conclusion in section 6.

II. LITERATURE SURVEY

A ton of examination has been done somewhat recently on voice based gender recognition. Various models are trained using Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Decision Trees(DT), Random Forests, gradient boosting, and Artificial Neural Networks are among the AI techniques (ANN).

[1]In the paper, "Comparative Study of Machine Learning Algorithms for Voice based Gender Identification", the authors used 6 different machine learning algorithms. The algorithms include K-Nearest Neighbor(KNN), Decision Trees(DT), Random Forest(RF), and types of support vector machines. It is observed the support vector machine gains higher accuracy 98.48% on the test data for the classification of gender.

[2]In the paper "Voice Gender Recognizer Recognition of Gender from Voice using Deep Neural Networks" by L.Jasuja, A. Rasool and G.Hajela. The authors proposed a multilayer perceptron which in tested on 3,168 records dataset and obtains accuracy of 96 percent.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

III. SOFTWARE LIBRARIES

A. *Python*

Python is a high-level, interpreted programming language known for its simplicity and ease of use. It has a large standard library and a thriving community of developers creating libraries and tools for a wide range of applications.

B. *Keras*

Keras is an open-source neural network library written in Python that provides a high-level API for building and training deep learning models.

C. *NumPy*

Numpy is a python library for scientific computing that provides support for multi-dimensional arrays, mathematical functions, and tools for working with arrays efficiently.

D. *Flask*

Flask is a lightweight, open-source web application framework for python that provides tools for building web applications quickly and efficiently, with support for extensions to add additional functionality.

E. *warbleR*

warbleR is an R package for the analysis of animal acoustic signals. It provides tools for visualizing and measuring acoustic parameters, as well as for automating analysis workflows.

F. *Rpy2*

Rpy2 is a python library that enables seamless integration between Python and R, allowing users to run R code from within Python and exchange data between the two languages. It provides a bridge for calling R functions from Python and vice versa.

IV. METHODOLOGY

A multilayer perceptron neural network model for voice-based gender classification is developed here. By using input as voice sample in .WAV format from the user, this model will be implemented on the web application for real-time classification of gender. The whole proposed process is depicted in fig1.

The following steps are carried out in this process:

1) The first step is to record the voice sample from the user for 20 seconds.

2) Acoustic analysis is performed on the recorded voice sample by using warbleR r package and storing the attributes along with the label of user in a .CSV file.

3) Building sequential multilayer perceptron model for the classification of gender. The model is trained by selecting random batches from dataset.

4) An Interactive web page is developed using flask and the trained model will be deployed on the web application where the input voice sample is taken from the user to classify the gender of the speaker.

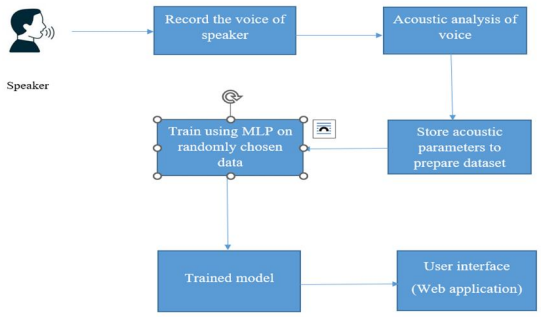


Fig 1: Block diagram of the proposed system

A. Recording of Voice of the Speaker

Voice samples are recorded from male and female speakers for 20 seconds. The audio format for the recordings used is .WAV files. Recorder.js is used as an api to record voice in .WAV file format. It consists startRecording() function which launches the promise based getUserMedia() and on success it passes the audio stream to an Audiocontext which is passed to our Recorder.js object and using timeout method the recorder automatically stops recording after 20seconds and creates a link to download the file.

B. Acoustic Analysis

The specan function from warbleR R package measures 22 acoustic parameters on the acoustic signals. The measured 22 parameters are in "Table I".

Table I: Measured Acoustic Properties

ACOUSTIC PROPERTIES	
PROPERTIES	DESCRIPTION
duration	length of the signal
meanfreq	mean frequency(in kHz)
sd	standard deviation of frequency
median	median frequency(in kHz)
Q25	first quantile(in kHz)
Q75	third quantile(in kHz)
IQR	interquantile range(in kHz)
skew	skewness
kurt	kurtosis
sp.ent	spectral entropy
sfm	spectral flatness
mode	mode frequency
centroid	frequency centroid
peakf	peak frequency
meanfun	average of fundamental frequency measured across acoustic signal
minfun	minimum of fundamental frequency measured across acoustic signal
maxfun	maximum of fundamental frequency measured across acoustic signal
meandom	average of dominant frequency measured across acoustic signal
mindom	minimum of dominant frequency measured across acoustic signal
maxdom	maximum of dominant frequency measured across acoustic signal
dfrange	range of dominant frequency measured across acoustic signal
modindx	modulation index

The acoustic parameters thus obtained are stored along with the label of classification as male or female into .CSV file.

C. Model Building Using MLP

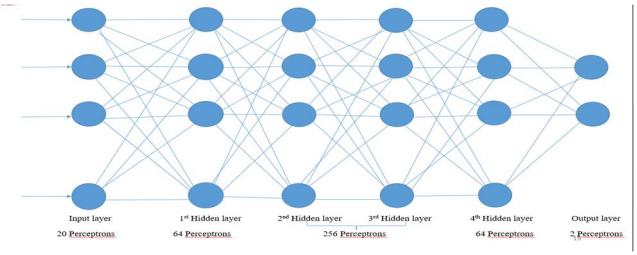


Fig 2: MLP architecture

All the training and test codes are written by using python libraries. Dataset has been loaded into python 2 dimensional array. Each row has 20 parameters and 1 label. The last column of data, which is label has been converted into 0 for male and 1 for female.
1 input layer, 4 hidden layers and 1 output layer have been used to build our model. Input layer have 20 inputs and has been connected to the first hidden layer which has 64 perceptrons. Second and third layers have each 256 perceptrons. The fourth hidden layer has 64 perceptrons which is connected to the output layer which has 2 perceptrons. Softmax activation function is used at the output layer to obtain the categorical distribution of the result for label. ReLU activation function has been used all hidden layers. Dropout 0.25 has been applied between each hidden layers.
Nadam optimizer in keras has been used to optimize the weights of the model while training of the model. The learning is chosen to be 0.0001 by choosing lower learning rate our model has been trained with 150 epochs.
5-fold cross validation has been used and average score has been obtained. Training and testing loops have been run 5 times. On each loop 20% of data has been used for testing and 10% of data has been used for validation.
Model weights has been stored to HDF5 file.

D. ReLU activation function

The rectified linear activation function, often known as ReLU, is a non-linear or piecewise linear function that, if the input is positive, outputs the input directly; if not, it outputs zero.
It is mathematically expressed as f(x) = max (0, x).

E. Softmax activation function

The outputs are normalized using the softmax function, which turns weighted sum values into probabilities that add up to one. Each value in the softmax function's output is regarded as the likelihood that a given class would contain that value.

F. NADAM Optimizer

Nadam optimizer is an adaptive optimization algorithm that combines the benefits of Adam optimizer and Nesterov momentum. It uses the gradient and momentum to update the learning rate, which helps to converge faster and avoid getting stuck in local minima.

G. System Architecture

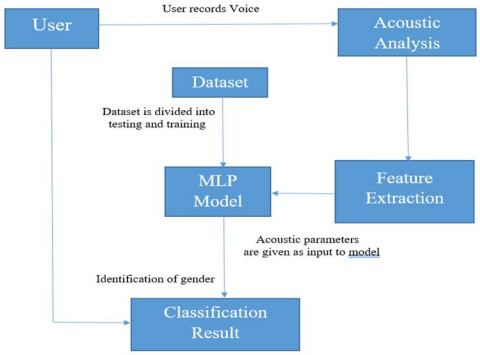


Fig 3: System architecture

H. Developing Interactive Webpage for gender classification

Flask is a web application framework for python that provides tools for building web applications quickly and efficiently. It provides simple interface for handling HTTP requests and responses, and supports extensions that can add additional functionality such a database integration and authentication.

The figure 4 shows the home page of the website.

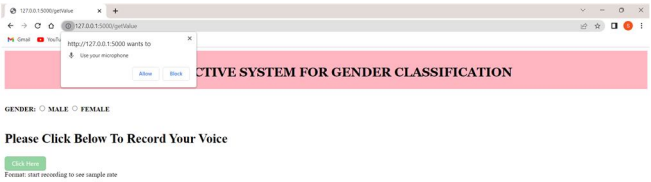


Fig 4: Home page

User clicks on click here button to record the voice and on click of allow microphone access the recorder gets started and records what user speaks for next 20 seconds and automatically stops and stores the recorded .WAV file in the local system of the user in the downloads folder. Rpy2 package helps executing r code and gives the downloaded file path to the specan function which measures the acoustic parameters after elimination of parameters not required the data is given to the trained model which is loaded and tested on various models such as MLP, KNN, SVM and finally displays results to the user.

The records of parameters and results are stored into csv file for future work.

V. RESULT AND EXECUTION

A 97% accuracy rate was achieved by the proposed model using 150 epochs while training.

The image in figure 5 shows the execution and results obtained.

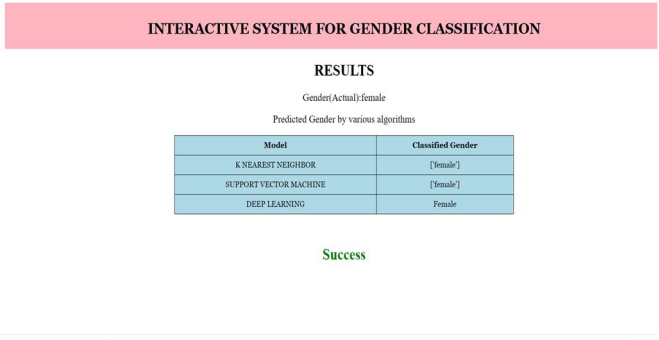


Fig 5: Output

Accuracy of different algorithms on dataset(3,168 records) while testing is noted down in “Table II”.

81 male and 44 female total [9]125 samples have been tested for KNN, SVM, MLP models and MLP has comparitvely gave less errors, accuracy obtained is nearly 84% on test data and results are noted down in “Table III”.

Table II: Result of accuracy on test data of various algorithms

Algorithm	Accuracy
K-Nearest Neighbor	95.2
Support vector machine	96
Multi-layer perceptron	97

Table III: Results on [9]Test Data on real time collected samples

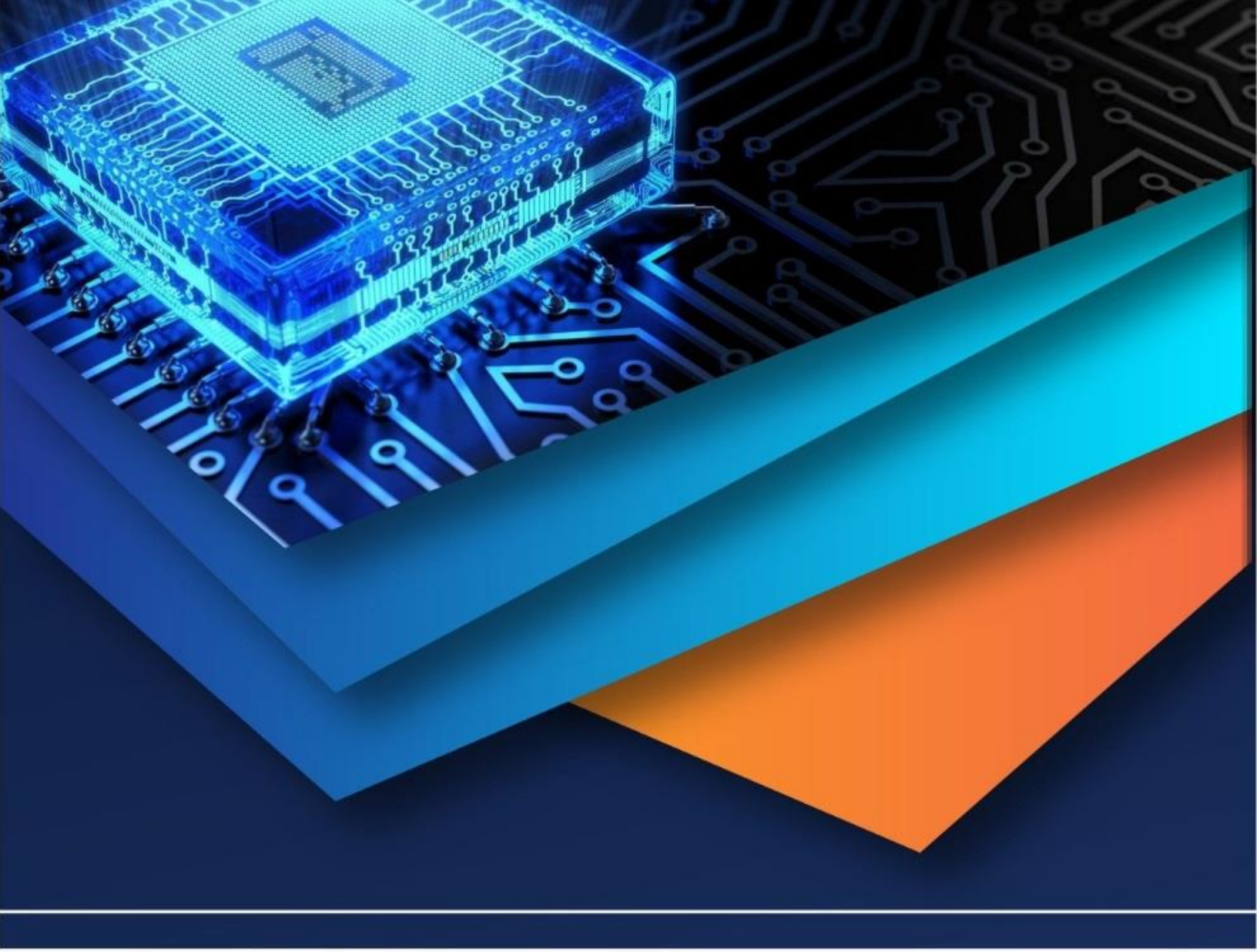
	Total No. of Tuples	KNN	SVM	MLP
Male	81	27	0	62
Female	44	40	44	44
Total Accuracy		53.6%	35.2%	84.8%

VI. CONCLUSION

This research provides a reliable and efficient method for classifying gender of the speaker using voice acoustic properties. For the detection of gender, Multilayer Perceptron neural network methods are applied. This System is very much beneficial in detecting criminals. It can also be used for marketing to find out the customer taste and preferences and increase the efficiency of the system. 125 real time voice samples have been tested by the webpage developed.

REFERENCES

[1] S. Srivastava, H. Sharma and D. Garg, "Comparative Study of Machine Learning Algorithms for Voice based Gender Identification," 2022 International Conference on Edge Computing and Applications (ICECAA), Tamilnadu, India, 2022, pp. 1136-1141, doi:10.1109/ICECAA55415.2022.9936549.
[2] L. Jasuja, A. Rasool and G. Hajela, "Voice Gender Recognizer Recognition of Gender from Voice using Deep Neural Networks," 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2020, pp. 319-324, doi: 10.1109/ICOSEC49089.2020.9215254.
[3] Khanum, Seema, and Marpe Sora. "Speech based gender identification using feed forward neural networks." Int. J. Comput. Appl (2015): 0975-8887.
[4] M. Alsulaiman, Z. Ali and G. Muhammad, "Gender Classification with Voice Intensity," 2011 UKSim 5th European Symposium on Computer Modeling and Simulation, Madrid, Spain, 2011, pp. 205-209, doi: 10.1109/EMS.2011.37.
[5] S. Chaudhary and D. K. Sharma, "Gender Identification based on Voice Signal Characteristics," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, 2018, pp. 869-874, doi: 10.1109/ICACCCN.2018.8748676.
[6] J. Shen, O. Lederman, J. Cao, F. Berg, S. Tang and A. Pentland, "GINA: Group Gender Identification Using Privacy-Sensitive Audio Data," 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 2018, pp. 457-466, doi: 10.1109/ICDM.2018.00061.
[7] S. M. S. I. Badhon, M. H. Rahaman and F. R. Rupon, "A Machine Learning Approach to Automating Bengali Voice Based Gender Classification," 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART), Moradabad, India, 2019, pp. 55-61, doi: 10.1109/SMART46866.2019.9117385.
[8] Dataset, [online] Available: https://www.kaggle.com/primaryobjects/voicegender.
[9] Dataset collected in real time (125 samples), [online] Available: https://github.com/Sirisha741/gender-classification/blob/bae2918ac156c6bfc20fb89c957955e4684bec6f/RealTimeTestData.csv
[10] M. Grimaldi and F. Cummins, "Speaker identification using instantaneous frequencies," IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 6, pp. 1097–1111, 2008
[11] K. Becker, "Identifying the Gender of a Voice using Machine Learning", 2016, unpublished.
[12] M. Li, K. J. Han, and S. Narayanan, "Automatic speaker age and gender recognition using acoustic and prosodic level information fusion," Computer Speech & Language, vol. 27, no. 1, pp. 151–167, 2013.
[13] Z. Li and D. Hoiem, "Learning without forgetting," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 12, pp. 2935–2947, 2018
[14] Python, [online] Available: https://docs.python.org/3/faq/general.html.
[15] R Wable, M. Araya-Salas and G. Smith-Vidaurre, "warbleR: an r package to streamline analysis of animal acoustic signals", Methods Ecol Evol, vol. 8, pp. 184-191, 2017.
[16] Num Py and E Travis, "Oliphant", A guide to NumPy, 2006, [online] Available: http://www.numpy.org.
[17] Tensorboard, Apr. 2020, [online] Available: https://www.tensorflow.org/tensorboard.
[18] JavaScript, [online] Available: https://www.w3schools.com/js/
[19] CSS, [online] Available: https://www.w3schools.com/css/
[20] By Armin Ronacher, "Flask Web Development, One Drop At A Time", http://flask.pocoo.org/,March 20 th 2015.
[21] Flask, [online] Available: https://www.mygreatlearning.com/blog/everything-you-need-to-know-about-flask-for-beginners/
[22] https://pythonprogramming.net/jinja-template-flask-tutorial/
[23] http://deeplearning.net/tutorial/_sources/mlp.txt
[24] Alsmadi, Mutasem & Omar, Khairuddin & Mohd Noah, Shahrul Azman. (2009). Back Propagation Algorithm: The Best Algorithm Among the Multi-layer Perceptron Algorithm. International Journal of Computer Science and Network Security. 9. 378-383.
[25] S. Haykin, Neural Networks: A Comprehensive Foundation (2 ed.). Prentice Hall, 1998.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)