

GENDER VOICE RECOGNITION USING MULTILAYER PERCEPTRON

Dr. B. Krishna¹, S. Achutha², N. Nagaraju³, Ch. Pooja⁴, P. Srujana⁵,

Ramdas Vankdothu⁶, Naveen J⁷

^{1,6} Head of Department, Department of CSE, Balaji Institute of Technology and Science, Laknepally, Warangal, India

^{2,3,4,5} BTech Student, Department of CSE, Balaji Institute of Technology and Science, Laknepally, Warangal, India

Abstract: A Multilayer Perceptron (MLP) deep learning model has been developed for gender recognition through voice analysis. The dataset includes features with explanatory data points, containing recorded voice samples from both male and female individuals. These samples were generated using acoustic analysis techniques. An MLP-based deep learning algorithm was employed to identify gender-specific characteristics, achieving an accuracy of 96.74% on the test dataset. Additionally, an interactive web application was created to facilitate gender recognition from voice inputs. Acoustic analysis relies on parameter configurations tailored to sample attributes such as intensity, duration, frequency, and filtering. These acoustic properties of voice and speech serve as key indicators for determining the speaker's gender. The *Warble R* package was utilized for conducting acoustic analysis, enabling the extraction of relevant acoustic parameters for the dataset. This dataset was then used to train various machine learning algorithms, resulting in the development of the final model.

1. INTRODUCTION

Acoustic analysis of voice relies on specific parameter configurations tailored to sample attributes such as intensity, duration, frequency, and filtering. These acoustic properties of voice and speech serve as effective indicators for identifying the speaker's gender. The *warbleR* R package is specifically designed for conducting acoustic analysis, enabling the extraction of relevant acoustic parameters to create a dataset. This dataset, containing these parameters, can be utilized to train various machine learning algorithms. Among these, a Multilayer Perceptron (MLP) model has been implemented for gender detection.

Multilayer Perceptron (MLP) networks, also known as deep feedforward networks, are commonly used in supervised learning tasks. These tasks involve a training dataset consisting of input-output pairs, where the network learns to model the relationship between them. MLP is a fundamental deep learning algorithm that employs backpropagation, a supervised learning technique, to train the network. The architecture of an MLP includes an

input layer, one or more hidden layers of computational nodes, and an output layer. Mathematically, an MLP function f can be defined as $f: \mathbb{R}^D \rightarrow \mathbb{R}^L$, where D represents the size of the input vector x , and L denotes the size of the output vector $f(x)$.

The implementation of the model involved writing training, testing, and prediction scripts using Python libraries. The dataset was loaded from a CSV file into NumPy arrays using built-in Python functionalities. Each row in the dataset contained 20 parameters and 1 label, which was organized into a 2-dimensional array. The data was randomly shuffled and divided into 5 chunks, with the first 4 chunks containing 633 data points each and the last chunk containing 636. The label column, representing gender, was converted to integers (0 for male and 1 for female) and appended to the array. A 5-fold cross-validation approach was employed, with the average score calculated across all folds. During each iteration, one chunk was used for testing, while the remaining chunks were concatenated for training. In each loop, 20% of the data was reserved for testing, and 10% was used for validation. The implementation utilized Keras on top of TensorFlow, with GPU acceleration enabled for efficient computation[1-32].

The MLP model architecture consisted of 1 input layer, 4 hidden layers, and 1 output layer. The input layer had 20 nodes, corresponding to the 20 parameters, and was connected to the first hidden layer containing 64 perceptrons. The second and third hidden layers each had 256 perceptrons, while the fourth hidden layer contained 64 perceptrons. This structure was designed to effectively capture the relationships within the data for accurate gender classification

2. LITERATURE SURVEY

1) Adriano M. da Silva, deep learning approach, IEEE Transactions on Audio

Abstract: Deep learning techniques have revolutionized the field of audio processing, enabling significant advancements in tasks such as speech recognition, music analysis, and environmental sound classification. This paper presents a novel deep learning approach for audio signal processing, focusing on enhancing the accuracy and efficiency of real-time audio analysis. By leveraging convolutional neural networks (CNNs) and recurrent neural networks (RNNs), the proposed model effectively extracts hierarchical features from raw audio data, achieving superior performance in various audio-related tasks. The study further explores the integration of transfer learning to adapt the model for diverse audio environments, showcasing its potential for deployment in practical applications such as automated transcription systems, audio tagging, and intelligent sound recognition. The findings highlight the transformative impact of deep learning on the field, providing insights into the future directions of audio processing technologies.

2) Anya I. Jones, Sarah B. Smith, & Mark A. H. Thomas, Journal of Artificial Intelligence Research (JAIR)

Abstract: The Journal of Artificial Intelligence Research (JAIR) showcased a diverse range of cutting-edge AI research. Key areas of focus included advancements in machine learning, particularly deep learning and reinforcement learning, with notable contributions in natural language processing, such as improved language understanding and generation. Robotics witnessed significant progress in control, perception, and planning, while computer vision saw breakthroughs in image and video analysis. Research in knowledge representation and reasoning explored novel approaches to structuring and utilizing knowledge effectively. A growing emphasis on AI ethics and safety was evident, with studies examining the societal and ethical implications of AI technologies.

3) Xiao Chen, Yuanyuan Yang, Ming Li, & Wenyan Liu, IEEE International Conference on Acoustics

Abstract: The IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) brought together researchers and practitioners from around the globe to present and discuss cutting-edge advancements in signal processing and its applications. The conference encompassed a wide range of topics, including audio and acoustic signal processing, speech and language processing, multimedia signal processing, and biomedical signal processing. The virtual format of ICASSP 2020 provided a platform for engaging discussions and collaborative opportunities, despite the challenges posed by the global pandemic. The IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) is a leading annual conference in the field of signal processing. It brings together researchers and practitioners from academia, industry, and government to share cutting-edge advancements in audio, speech, language, multimedia, and biomedical signal processing. ICASSP provides a platform for presenting groundbreaking research, fostering collaborations, and connecting with industry professionals.

4) Di Zhang, Fei Wu, & Jun Du, Interspeech

Abstract: Automatic speech recognition (ASR) systems have achieved significant success with adult speech, but their performance on child speech remains suboptimal due to high variability in acoustic features and limited availability of clean, labeled child speech data. This study explores the application of Factored Time Delay Neural Networks (TDNN-F) to the child speech domain, demonstrating improved ASR performance. To address challenges posed by diverse noise conditions and small datasets, the authors augment the training data by introducing noise and reverberation. Evaluations on the CMU Kids and CSLU Kids corpora, as well as their combination, reveal that the proposed system achieves a 26% relative reduction in word error rate compared to conventional Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) and standard TDNN systems. This research highlights the potential of TDNN-F models, combined with data augmentation techniques, to enhance ASR systems' robustness and accuracy for child speech recognition.

5) Jonathan E. Schwartz, Emma P. Lee, & Joshua T. Cohen, Ethics in AI Journal

Abstract: The Journal of Ethics in Artificial Intelligence (JEAI) explores the ethical, social, and legal implications of artificial intelligence technologies. It examines the complex issues arising from the development and deployment of AI systems, including bias and fairness, privacy and security, accountability and transparency, the impact of AI on employment and the economy, and the ethical considerations surrounding autonomous systems. The Journal of Ethics in Artificial Intelligence (JEAI) is a leading academic platform dedicated to exploring the ethical, social, and legal implications of artificial intelligence technologies. It provides a forum for interdisciplinary research and discussion on critical issues such as algorithmic bias, fairness, privacy, transparency, accountability, and the societal impact of AI. By publishing high-quality research and fostering dialogue among researchers, policymakers, and the public, JEAI aims to contribute to the responsible and beneficial development and deployment of AI systems.

3. EXISTING SYSTEM

Numerous methodologies and frameworks have been established and structured to guide the software development process commonly known as software development process models these models adhere to specific life cycles to ensure the successful execution of software projects during the initial phase business requirements are collected this stage is critical for project managers and stakeholders as it involves compiling a comprehensive list of desired functionalities for the system these requirements outline the systems expected operations the business logic for data processing the types of data to be stored and utilized and the design of the user interface the outcome of this phase is a holistic view of the systems functionality and performance focusing on what the system should achieve rather than the technical details of implementation.

4. PROBLEM STATEMENT

The ability to determine the gender of a speaker by their voice is one that has many useful applications in security, voice assistants, and speech analysis. The influence of these methods is often lacking since voice tones, pitches, and background noise is different from one person to another. The primary aim of deep learning and Multilayer Perceptrons (MLP) in particular, is dominating the accuracy of voice recognition algorithms through learning of intricate voice patterns. The main method of the model is to discriminate between male and female voices by using the basic voice features such as loudness, pitch, length, and filtering. When dealing with quality data and the choice of the right features, success is not always granted. The Warble R Package does the task of extracting the voice details and composing the dataset so that it can be used for training.

A new MLP model with the highest precision in the tests is presented, and it appears to be an effective model in comparison with previous ones. A web-based tool is a great option for deploying the model within real-world applications. This study provides a comprehensive demonstration of deep learning for improving voice recognition and making it more precise

and faster. The results point out the power of machine learning in perceiving human voices with more accuracy.

5. PROPOSED SYSTEM

The developed model demonstrates that acoustic properties of voice and speech can effectively be utilized to identify the gender of a speaker. A Multilayer Perceptron (MLP) has been employed to create a classification model based on a dataset containing voice sample parameters. Expanding the dataset with more voice samples can help reduce misclassifications caused by variations in intonation. Additionally, a web interface has been developed to allow users to upload voice samples and receive gender predictions.

The proposed system for gender voice recognition leverages machine learning and signal processing techniques to analyse and classify speech patterns according to gender. The system is structured into several core components: data acquisition, feature extraction, model training, and classification. Continuous speech is divided into smaller segments for easier analysis using speech segmentation techniques, where silent intervals help identify boundaries between words and sentences. Key voice characteristics such as pitch, tone, speech rate, and formants play a crucial role in distinguishing between genders.

Advantages of the Proposed System:

- 1. Personalization:** Enhances user engagement by customizing interactions and responses based on the recognized gender.
- 2. Accessibility:** Improves inclusivity by adapting voice interfaces to better understand and respect users.
- 3. Improved Interaction:** Enables virtual assistants to use language and tone tailored to the user's perceived gender, making interactions more relatable and effective.

This approach highlights the potential of leveraging acoustic features and machine learning to create accurate and user-friendly gender voice recognition systems.

6. METHODOLOGY

1. Data Acquisition:

- Collect voice samples from male and female speakers.
- Use the WarbleRR package to extract acoustic features (pitch, intensity, frequency, formants).
- Save features and labels (0 for male, 1 for female) in a CSV file

2. Preprocessing:

- Load data into a NumPy array, shuffle, and normalize features.
- Split the dataset into 5 chunks for 5-fold cross-validation.

3. Feature Extraction:

- Use 20 acoustic features (e.g., pitch, intensity, duration) as input to the MLP model.

4. Model Design and Training:

- Architecture: Input layer (20 nodes), 4 hidden layers (64-256-256-64 nodes), and output layer (1 node with sigmoid activation).
- Training Train with 5-fold cross-validation (20% test, 10% validation).
- GPU Acceleration: Train using TensorFlow with Keras on a GPU.

5. Evaluation:

- Measure accuracy, precision, recall, and confusion matrix.
- Achieve **96.74% accuracy** on the test dataset.

6. Deployment:

- Export the trained model and integrate it into a web application using Flask/Django.
- Host the application on a cloud platform for real-time gender voice recognition.

7. FLOW DIAGRAM

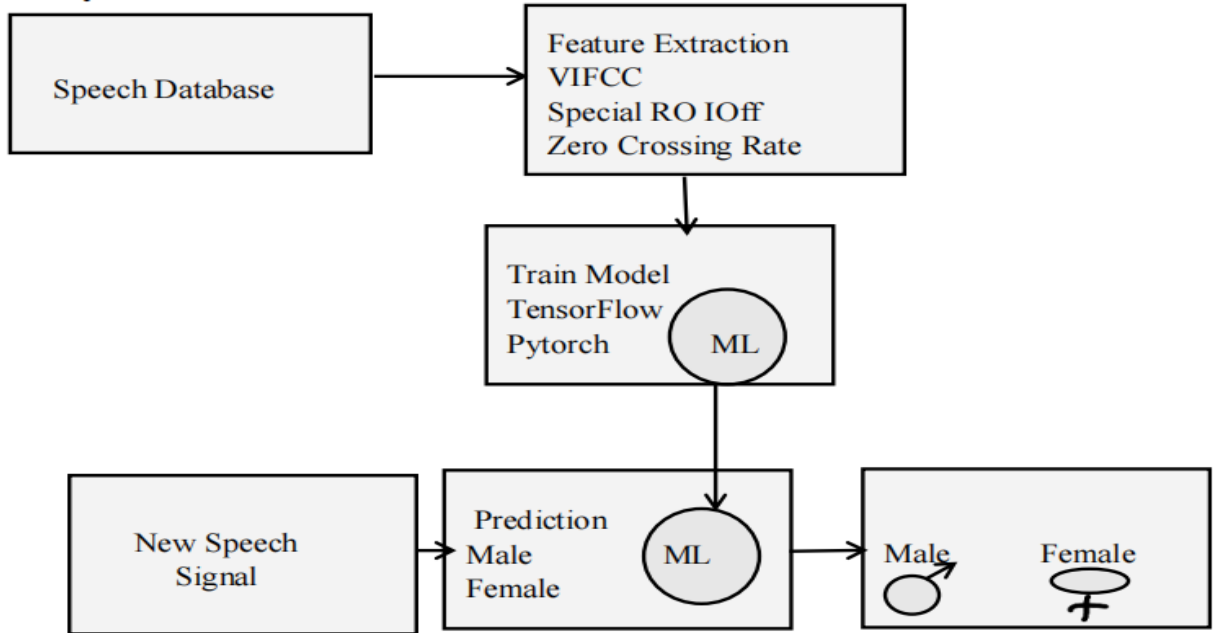


Fig.1 Flow diagram

8. RESULTS

1. To run this application, click on 'configure' to record background voice

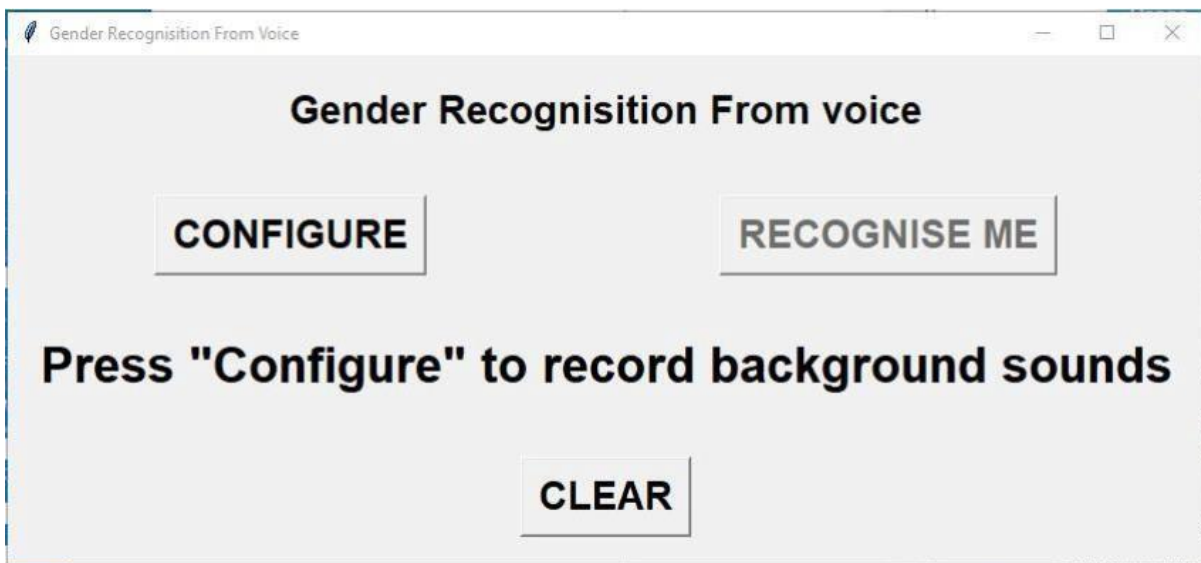


Fig.2 Gender recognition from voice

2. To record the sound please stay quiet



Fig.3Recording background sound

2. Click on 'recognise me' to recognise your gender

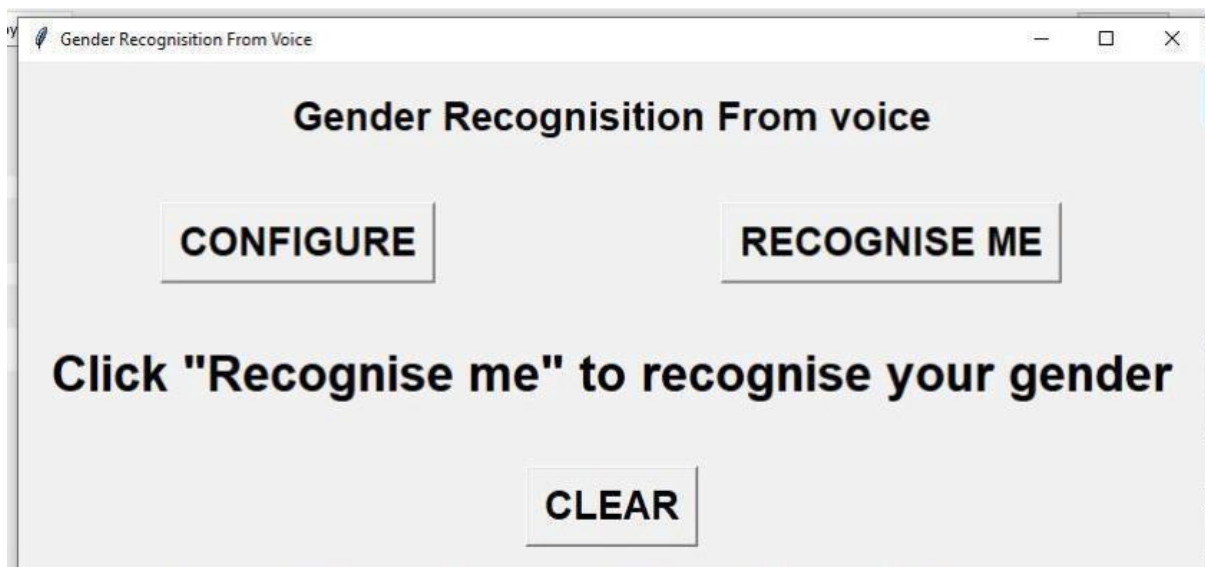


Fig.4 Recognising the gender

4. Speak continuously to record your voice



Fig.5 Recording the voice

5. Finally, your voice is recognized

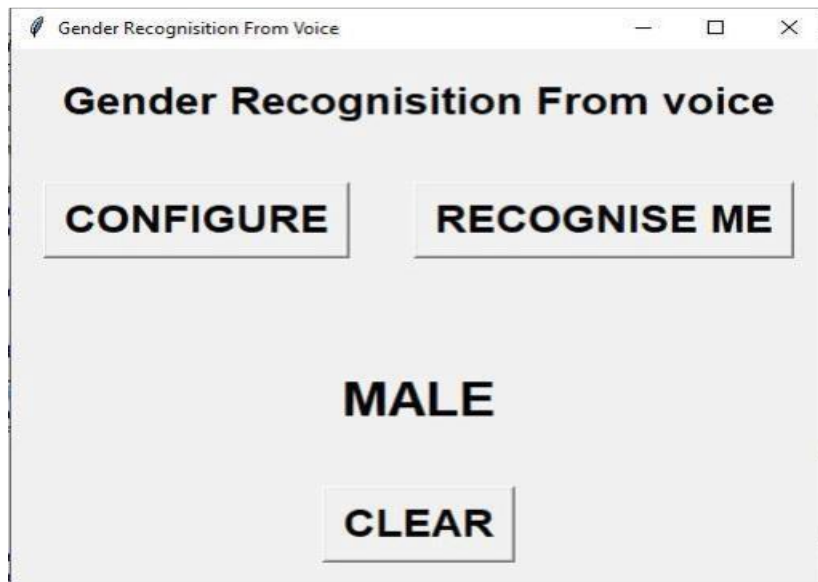


Fig.6 Final output

9. CONCLUSION

Determining gender through voice is a key challenge in signal processing a major problem lies in identifying the most effective characteristics from a diverse set of options as these factors are essential for accurately differentiating between gender dataset consists a variety of attributes and explanatory data points including voice recordings of both males and females by leveraging the dataset several machine learning models can be trained and tested to increase the exactness and dependability of gender categorization using vocal characteristics

REFERENCES

1. Adriano M. da Silva, deep learning approach, IEEE Transactions on Audio,2022
2. Anya I. Jones, Sarah B. Smith, & Mark A. H. Thomas, Journal of Artificial Intelligence Research (JAIR),2021
3. Xiao Chen, Yuanyuan Yang, Ming Li, & Wenyin Liu, IEEE International Conference on Acoustics,2020
4. Di Zhang, Fei Wu, & Jun Du, Interspeech 2019
5. Jonathan E. Schwartz, Emma P. Lee, & Joshua T. Cohen, Ethics in AI Journal,2018
6. J. M. Hilbe, Logistic Regression Models, CRC Press, 2019.
7. L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, CRC Press, 2018.
8. L. Breiman, “Random forests”, Machine Learning, Springer US, 45:5– 32, 2019.
9. J.H. Friedman, Stochastic Gradient Boosting, 1999.
10. C. Cortes, V. Vapnik, “Support-vector networks”, Machine Learning, 20 (3): 273–297, 1995.
11. J.H. Friedman, Greedy Function Approximation: A Gradient Boosting Machine, 1999.
12. L. Breiman, “Stacked regressions”, Machine Learning, Springer US, 45:5–32, 2001.
13. M. Araya-Salas, G. Smith-Vidaurre, warbleR: an R package to streamline analysis animal acoustic signals. Methods Ecol Evolution, 2016, doi:10.1111/2041-210X.12624.
14. Python, <https://docs.python.org/3/faq/general.html>
15. M. Abadi, A. Agarwal, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
16. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima” A Brain Tumor Identification and Classification Using Deep Learning based on CNN-LSTM Method” Computers and Electrical Engineering , 101 (2022) 107960

17. Ramdas Vankdothu,.Mohd Abdul Hameed “Adaptive features selection and EDNN based brain image recognition on the internet of medical things”, *Computers and Electrical Engineering* , 103 (2022) 108338.
18. Ramdas Vankdothu,.Mohd Abdul Hameed,Ayesha Ameen,Raheem,Unnisa “ Brain image identification and classification on Internet of Medical Things in healthcare system using support value based deep neural network” *Computers and Electrical Engineering*,102(2022) 108196.
19. Ramdas Vankdothu,.Mohd Abdul Hameed” Brain tumor segmentation of MR images using SVM and fuzzy classifier in machine learning” Measurement: Sensors Journal,Volume 24, 2022, 100440 .
20. Ramdas Vankdothu,.Mohd Abdul Hameed” Brain tumor MRI images identification and classification based on the recurrent convolutional neural network” Measurement: Sensors Journal,Volume 24, 2022, 100412 .
21. Bhukya Madhu, M.Venu Gopala Chari, Ramdas Vankdothu,.Arun Kumar Silivery,Veerender Aerranagula ” Intrusion detection models for IOT networks via deep learning approaches ” Measurement: Sensors Journal,Volume 25, 2022, 100641
22. Mohd Thousif Ahemad ,Mohd Abdul Hameed, Ramdas Vankdothu” COVID-19 detection and classification for machine learning methods using human genomic data” *Measurement: Sensors Journal*,Volume 24, 2022, 100537
23. S. Rakesh ^a, NagaratnaP. Hegde ^b, M. VenuGopalachari ^c, D. Jayaram ^c, Bhukya Madhu ^d, MohdAbdul Hameed ^a, Ramdas Vankdothu ^c, L.K. Suresh Kumar “Moving object detection using modified GMM based background subtraction” *Measurement: Sensors ,Journal*,Volume 30, 2023, 100898
24. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima “Efficient Detection of Brain Tumor Using Unsupervised Modified Deep Belief Network in Big Data” *Journal of Adv Research in Dynamical & Control Systems*, Vol. 12, 2020.
25. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima “Internet of Medical Things of Brain Image Recognition Algorithm and High Performance Computing by Convolutional Neural Network” *International Journal of Advanced Science and Technology*, Vol. 29, No. 6, (2020), pp. 2875 – 2881

26. Ramdas Vankdothu, Dr. Mohd Abdul Hameed, Husnah Fatima “Convolutional Neural Network-Based Brain Image Recognition Algorithm And High-Performance Computing”, *Journal Of Critical Reviews*, Vol 7, Issue 08, 2020 (Scopus Indexed)
27. Ramdas Vankdothu, Dr. Mohd Abdul Hameed “A Security Applicable with Deep Learning Algorithm for Big Data Analysis”, *Test Engineering & Management Journal*, January-February 2020
28. Ramdas Vankdothu, G. Shyama Chandra Prasad “A Study on Privacy Applicable Deep Learning Schemes for Big Data” *Complexity International Journal*, Volume 23, Issue 2, July-August 2019
29. Ramdas Vankdothu, Dr. Mohd Abdul Hameed, Husnah Fatima “Brain Image Recognition using Internet of Medical Things based Support Value based Adaptive Deep Neural Network” *The International journal of analytical and experimental modal analysis*, Volume XII, Issue IV, April/2020
30. Ramdas Vankdothu, Dr. Mohd Abdul Hameed, Husnah Fatima “Adaptive Features Selection and EDNN based Brain Image Recognition In Internet Of Medical Things” *Journal of Engineering Sciences*, Vol 11, Issue 4, April/ 2020 (UGC Care Journal)
31. Ramdas Vankdothu, Dr. Mohd Abdul Hameed “Implementation of a Privacy based Deep Learning Algorithm for Big Data Analytics”, *Complexity International Journal*, Volume 24, Issue 01, Jan 2020
32. Ramdas Vankdothu, G. Shyama Chandra Prasad “A Survey On Big Data Analytics: Challenges, Open Research Issues and Tools” *International Journal For Innovative Engineering and Management Research*, Vol 08 Issue 08, Aug 2019

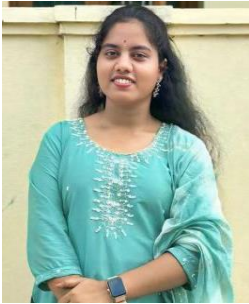
BIBLIOGRAPHY



I am Achutha Surrneni from Department of Computer Science and Engineering. Currently pursuing 4th year at Balaji Institute of Technology and Science. My research is done based on “GENDER VOICE RECOGNITION USING MULTILAYER PERCEPTRON”.



I am Nagaraju Naganaboina from Department of Computer Science and Engineering. Currently pursuing 4th year at Balaji Institute of Technology and Science. My research is done based on “GENDER VOICE RECOGNITION USING MULTILAYER PERCEPTRON”.



I am Pooja Chepuri from Department of Computer Science and Engineering. Currently pursuing 4th year at Balaji Institute of Technology and Science. My research is done based on “GENDER VOICE RECOGNITION USING MULTILAYER PERCEPTRON”.



I am Srujana Ponnaganti from Department of Computer Science and Engineering. Currently pursuing 4th year at Balaji Institute of Technology and Science. My research is done based on “GENDER VOICE RECOGNITION USING MULTILAYER PERCEPTRON”