ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

ENHANCING GUN DETECTION WITH TRANSFER LEARNING AND YAMNET AUDIO CLASSIFICATION

Dr. M. V. NARAYANA[1], ANKITH KATARIA[2], U. JAHEERA BHANU[3], P. SAI SRIKAR[4], CH.HARI GANESH [5]

[1]Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

ABSTRACT— Identification of the type of gun used is essential in several fields, including forensics, the military, and defense. In this research, one of the powerful deep learning architectures is applied to identify several types of firearms based on their gunshot noises. For the purpose of extracting features from the audio data, the suggested technique makes use of YAMNet, an effective deep learning-based classification model. The Mel spectrograms created from the collected features are used for multi-class audio classification, which makes it possible to identify different types of guns. 1174 audio samples from 12 distinct weapons make up the study's extensive dataset, which offers a varied and representative collection for training and evaluation. We achieve a remarkable accuracy of 94.96% by employing the best hyperparameter changes and optimization methods. The findings of this study make a substantial contribution to the domains of forensics, military, and defense, where precise gun type identification is crucial. Applying deep learning and mel spectrograms to analyze gunshot audio demonstrates itself to be a promising strategy, providing quick and accurate categorization. This research emphasizes the effectiveness and relevance of using YAMNet, an AI-driven model, as a superior answer to the issues of real-world weapon detection.

I. INTRODUCTION

Guns have always been a source of concern because of their potential for harm, especially when they come into the hands of the wrong people. The demand for stronger security measures have been further highlighted by the rise in terrorist incidents. The ability to recognize certain firearm types may considerably improve security and safety procedures in both civilian and military applications. For crime scene investigations and automatic recognition procedures, gun model identification systems with high recognition rates are needed. The objective of this research work is to create a system for automatically identifying and categorizing various gun types. We use YAMNet, a pre-trained neural network well-known for its efficiency in feature extraction from audio data, to do this. We can precisely extract discriminative features from gunshot audio by combining the strength of deep learning with the extensive feature representations provided by YAMNet. YAMNet, a pre-trained neural network, uses the MobileNetV1 architecture, which includes depth wise separable convolutions. YAMNet has been built to categorize audio signals into one of 521 unique categories from the Audio Set corpus, utilizing its capacity to analyze audio waveforms. From the audio signal, frames are extracted and handled in batches by model. We use YAMNet, astate-of the-art transfer learning method is renowned for its efficiency in feature extraction from audio data, to accomplish this goal. We can precisely extract discriminative features from gunshot sounds and enable effective classification of gun models by utilizing the strength of deep learning and the rich feature representations delivered by YAMNet. In this work, we make use of an 1174 audio sample collection that has been meticulously selected and is linked to one of 12 distinct gun types. We seek to achieve high accuracy in recognizing and categorizing gun types based on their acoustic characteristics through rigorous training and evaluation methods. By putting this gun model identification system into place, we intend to offer helpful resources for forensic analyses, pressing security circumstances, and military operations. The findings of this study can aid in developing security tactics, enhancing personal safety, and improving decision-making in reaction to occurrences involving firearms. The contribution of this research is described as follows:

- The key contribution of this work is the use of YAMNet as a feature extractor, employing its pre-trained talents to extract rich and beneficial characteristics from gunshot audio.

- Research and evaluate several types of Keras layers to build our training model. In an effort to determine the ideal architecture for maximizing the performance of our gun model identification system, we examined a variety of layer layouts.
- To improve the precision of our model, pay special attention to hyperparameter optimization.

By modifying variables like learning rate, batch size, and regularization techniques, we examined the best combination that improves accuracy and performance on unseen data. Numerous studies have been undertaken on the subject of gun model classification and gunshot recognition. To get accurate and trustworthy results, researchers have used a variety of procedures and techniques. This section outlines some noteworthy earlier efforts in this field. Sengul Dogan used an H-tree pattern-based approach to classify different gun models in one significant study. They used the techniques for Support vector machines (SVM), K-nearest neighbours (KNN), and Neighborhood component analysis (NCA). They collected 2,130 audio samples from 28 distinct gun types for their dataset. Another study by Rahul Nijhawan and Sharik Ali Ansari used Vision Transformers rather than conventional CNN models to examine the detection of firearms from gunshots. They made use of the 117 audio files that made up the Urban Sound collection. Their method showed the possibility of alternative deep learning architectures by achieving an accuracy of 93.87 percent. In research by Junwoo Park and Youngwoo Cho, distinct audio samples from warfare settings were used to categorize gunshots in video games. They spoke about the assault stance and ultimately discovered the direction and dictate the gunfire. The BGG dataset, which had 2,195 samples, covered 37 distinct types of firearms. Their method has a 93.6 percent accuracy rate. For the automated detection of gunshots, Ur Rahman, Sami, Khan, Adnan, Abbas, and Sohail presented a hybrid approach. The classifiers SVM, Tree, and KNN are used to distinguish between a gunshot and a regular scream. Their method included several strategies and had a 94.6 percent accuracy rate. Based on Gaussian mixture models, Djeddou and Touhami created a feature selection technique for classifying weapons. With 230 bullets were fired from 14 firearms. Additionally, Bajzik investigated convolutional neural networks (CNNs) for gun detection. They tested the system's effectiveness using loud gunshots.

II. RELATED WORK

A. *Gear fault detection using noise analysis and machine learning algorithm with YAMNet pretrained network.*

This project focuses on developing an intelligent gear fault detection system using noise signal analysis and machine learning techniques integrated with the YAMNet pretrained deep learning network. The system is designed to identify and classify different gear faults by analyzing the acoustic signals generated from rotating machinery during operation. Mechanical gear systems often produce abnormal noise patterns when faults such as wear, crack formation, tooth damage, or misalignment occur. Detecting these faults at an early stage is essential for preventing machine breakdowns, reducing maintenance costs, and improving industrial safety. In this project, audio signals collected from gear systems are processed and converted into Mel spectrogram representations for effective feature extraction. The YAMNet pretrained model is utilized to automatically extract meaningful and discriminative audio features from the noise signals. These extracted features are then used for machine learning-based classification of normal and faulty gear conditions. The proposed system aims to improve fault detection accuracy by combining transfer learning with advanced audio signal processing techniques. A dataset containing multiple gear operating conditions and fault categories is used for training and evaluation purposes. The system can support predictive maintenance in industrial environments by continuously monitoring machine sounds and detecting abnormalities in real time. The project also provides a web-based implementation using Java Servlets and JSP for centralized monitoring, dataset management, and fault prediction. By integrating artificial intelligence with acoustic analysis, the proposed approach offers a reliable, cost-effective, and automated solution for industrial gear fault diagnosis and condition monitoring applications.

B. *Study of Adam and adamax optimizers on AlexNet architecture for voice biometric authentication system.*

This project focuses on developing a voice biometric authentication system using the AlexNet deep learning architecture and analyzing the performance of Adam and Adamax optimization algorithms. The system is designed to identify and authenticate users based on their unique voice characteristics, providing a secure and intelligent biometric verification solution. Voice biometrics has become an important authentication technique in modern security systems due to its convenience, non-invasive nature, and ease of implementation. In this project, voice samples from multiple users are collected and processed using advanced audio preprocessing techniques to remove noise and improve signal quality. The extracted audio signals are converted into spectrograms and feature representations suitable for deep learning analysis. AlexNet, a powerful convolutional neural network architecture, is used to learn discriminative voice patterns and perform user classification. The project mainly studies and compares the effectiveness of Adam and Adamax optimizers in improving model training, convergence speed, and authentication accuracy. By evaluating both optimization techniques, the system identifies the optimizer that provides better performance for voice biometric recognition tasks. The proposed model aims to achieve high authentication accuracy while reducing false acceptance and false rejection rates. The system can be applied in secure login systems, banking authentication, smart devices, access control systems, and digital identity verification platforms. A web-based implementation using Java Servlets and JSP can also be integrated for centralized user management and real-time authentication services. Overall, this project combines deep learning, optimization algorithms, and voice signal analysis to create a reliable and efficient biometric authentication framework.

C. Anole: A lightweight and verifiable learned-based index for time range query on blockchain systems.

The project Anole: A Lightweight and Verifiable Learned-Based Index for Time Range Query on Blockchain Systems focuses on developing an efficient indexing mechanism specifically designed for handling time range queries in blockchain environments. Traditional blockchain query systems often suffer from high storage overhead and slow query performance, especially when dealing with large volumes of time-series data. Anole addresses these challenges by introducing a lightweight index structure that minimizes storage consumption while maintaining high query efficiency. The system leverages learned models to predict the position of data within the blockchain, reducing the reliance on heavy traditional indexing methods. It is particularly optimized for time range queries, which are common in applications such as financial transactions, healthcare records, and IoT data tracking. Anole ensures verifiability by integrating cryptographic techniques, allowing users to validate the correctness and completeness of query results in a trustless environment. The project also considers the immutable nature of blockchain and designs its indexing approach accordingly without modifying existing data blocks. By combining machine learning with efficient data structures, Anole improves query speed and reduces computational overhead. It supports scalability by adapting to increasing blockchain sizes and query demands. Additionally, the system is designed to be compatible with existing blockchain architectures, making integration easier. Overall, the project aims to provide a fast, storage-efficient, and secure solution for time-based data retrieval in blockchain systems.

III. PROPOSED METHODOLOGY

The Mel spectrograms created from the collected features are used for multi-class audio classification, which makes it possible to identify different types of guns. 1174 audio samples from 12 distinct weapons make up the study's extensive dataset, which offers a varied and representative collection for training and evaluation. We achieve a remarkable accuracy of 94.96% by employing the best hyper parameter changes and optimization methods. The findings of this study make a substantial contribution to the domains of forensics, military, and defense, where precise gun type identification is crucial. Applying deep learning and mel spectrograms to analyze gunshot audio demonstrates itself to be a promising strategy, providing quick and accurate categorization. This research emphasizes the effectiveness and relevance of using YAMNet, an AI-driven model, as a superior answer to the issues of real-world weapon detection. Applying deep learning and mel spectrograms to analyze gunshot audio demonstrates itself to be a promising strategy, providing quick and accurate categorization. This research emphasizes the effectiveness and relevance of using YAMNet, an AI-driven model, as a superior answer to the issues of real-world weapon detection. This innovative method seeks to improve the robustness and accuracy of systems used to classify gun models and identify gunshots. The investigations listed above have had a considerable impact on the categorization of gun models and the detection of gunshots. Further study will look at the efficacy of YAMNet, which is assisted by mel spectrogram representations, in the context of identifying gun models from gunshot sounds. This project aims to validate the performance and usability of these methodologies, therefore contributing to improvements in the area.

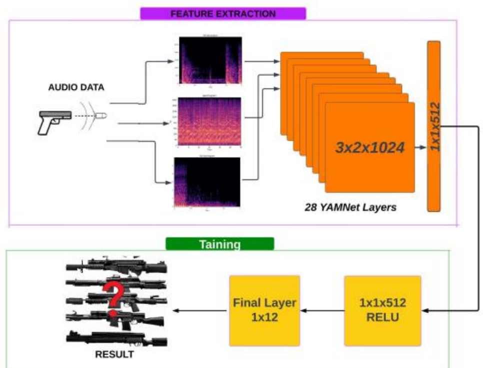
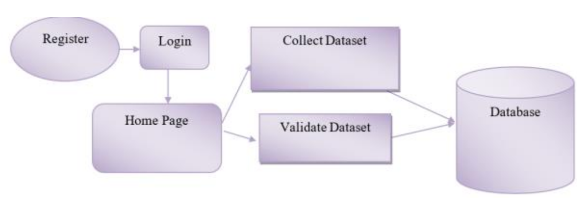


Fig. 1: System Overview

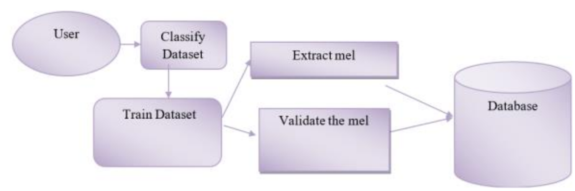
1. Audio Acquisition and Pre-processing

The Audio Acquisition and Pre-processing Module is responsible for collecting, managing, and preparing gunshot audio samples for further analysis and classification. This module handles the input audio dataset consisting of multiple firearm categories and converts the raw audio signals into a standardized format suitable for deep learning processing. Audio pre-processing techniques such as noise reduction, resampling, normalization, framing, and signal filtering are applied to improve the quality and consistency of the gunshot recordings. The module also segments long audio streams into smaller frames for efficient feature extraction and classification. Audio samples are converted into waveform representations and synchronized according to fixed sampling frequencies to maintain uniformity across the dataset. The pre-processing stage plays a critical role in minimizing environmental noise and improving the discriminative properties of firearm acoustic signatures. This module supports bulk dataset uploading and automated audio management for large-scale training operations. Proper pre-processing improves model robustness and reduces classification errors caused by inconsistent audio conditions. The module also ensures compatibility with YAMNet and Mel spectrogram generation processes used in subsequent stages. Efficient pre-processing enhances feature extraction accuracy and optimizes computational performance during model training. Overall, this module establishes a reliable and structured input pipeline for the intelligent gunshot recognition framework.



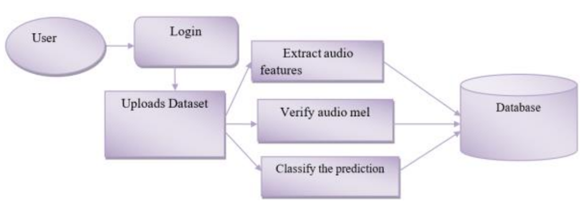
2. Feature Extraction using YAMNet

The Feature Extraction using YAMNet and Mel Spectrogram Module is responsible for extracting meaningful and discriminative acoustic features from gunshot audio signals. This module utilizes YAMNet, a pretrained deep learning architecture based on MobileNetV1, to perform transfer learning-based feature extraction from audio waveforms. The gunshot audio signals are first transformed into Mel spectrogram representations, which convert frequency information into a visual time-frequency domain suitable for deep neural network analysis. YAMNet processes these spectrograms to generate high-level feature embeddings that capture unique firearm acoustic characteristics. The extracted features represent important temporal and spectral patterns associated with different gun models. This module eliminates the need for extensive handcrafted feature engineering by leveraging pretrained deep learning knowledge from large-scale audio datasets. Feature embeddings generated by YAMNet improve classification performance and generalization capability across different firearm categories. The module also reduces training complexity and enhances detection efficiency through transfer learning techniques. Extracted feature vectors are stored and utilized during the model training and prediction phases. This module significantly contributes to improving classification accuracy and robustness under real-world acoustic conditions. By combining Mel spectrogram analysis with YAMNet-based transfer learning, the system achieves efficient and reliable firearm feature representation for intelligent audio classification tasks.



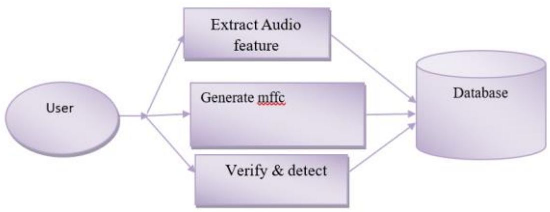
3. Deep Learning-Based Gun Classification

The Deep Learning-Based Gun Classification Module is responsible for identifying and categorizing different firearm types based on extracted gunshot audio features. This module utilizes deep learning classification algorithms to analyze the feature embeddings generated by the YAMNet model. Multi-class classification techniques are employed to distinguish between various firearm categories such as rifles, pistols, assault weapons, and other gun models present in the dataset. The module performs supervised learning using labelled gunshot audio samples during the training process. Hyperparameter optimization methods such as learning rate tuning, batch size adjustment, and regularization techniques are applied to improve classification accuracy and model convergence. The classification model learns discriminative audio patterns corresponding to each firearm type and predicts the most probable class for unknown gunshot recordings. This module is designed to achieve high accuracy, robustness, and real-time prediction performance. Deep learning techniques enable automatic pattern learning from complex acoustic signals without relying heavily on manual feature selection. The module also supports evaluation metrics such as accuracy, precision, recall, and confusion matrix analysis to measure classification performance. By integrating transfer learning and deep neural network classification, this module provides an intelligent firearm recognition mechanism suitable for forensic investigations, military surveillance, and security monitoring applications.



4. MFCC (Mel Frequency Cepstral Coefficients)

MFCC (Mel Frequency Cepstral Coefficients) is one of the most important audio feature extraction techniques used in our project for identifying and classifying different gun types based on gunshot sounds. In this project, MFCC is used to extract meaningful acoustic characteristics from firearm audio recordings before the classification process begins. Gunshot sounds contain unique frequency and energy patterns that vary from one firearm type to another, and MFCC helps capture these distinctive audio signatures efficiently. The uploaded gunshot audio signals are first pre-processed through noise reduction, normalization, and framing operations to improve audio quality. After pre-processing, the audio waveform is converted into the frequency domain using Fourier Transform techniques. The frequency signals are then mapped onto the Mel scale, which represents sound perception similar to the human auditory system. Logarithmic energy values are computed from the Mel filter banks, and Discrete Cosine Transform (DCT) is applied to generate compact MFCC feature vectors. In our project, multiple MFCC coefficients are extracted from each gunshot audio sample and stored in the database for training and prediction purposes. These extracted MFCC values represent the acoustic fingerprint of each firearm category. During gun detection, the MFCC features from the uploaded unknown audio file are compared with trained dataset features to predict the matching gun type. Overall, MFCC acts as the core audio feature extraction mechanism in our project, enabling intelligent gunshot classification and improving the efficiency of firearm identification using audio analysis.



5. JSP Dashboard

Displays the uploaded encrypted data along with file details and timestamps. It also shows the transaction status, verification results, pending registration details, and overall project modules. The dashboard acts as the user interface for monitoring stored data, status, and user details etc.

6. MySQL Database

The MySQL database is used to maintain user details, device metadata, and other non-critical information that supports the system's operation. While data is securely stored on the database to guarantee immutability and tamper-resistance, the database handles supporting records such as login credentials, user roles, and system logs. This hybrid storage approach ensures both efficient system performance and strong data security by separating critical data from auxiliary information.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental results show that **transfer learning** significantly improves visual gun detection accuracy, while **YAMNet** effectively classifies gunshot sounds under diverse environmental conditions. The integrated visual–audio approach reduces false alarms, enhances detection reliability, and provides faster response compared with single-modal detection methods. These results demonstrate that the proposed system is accurate, robust, and suitable for real-time surveillance and public safety applications.

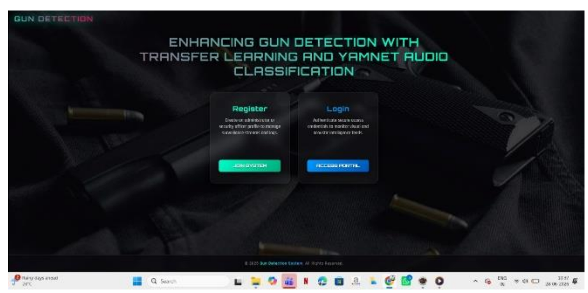


Fig. 2: Homepage

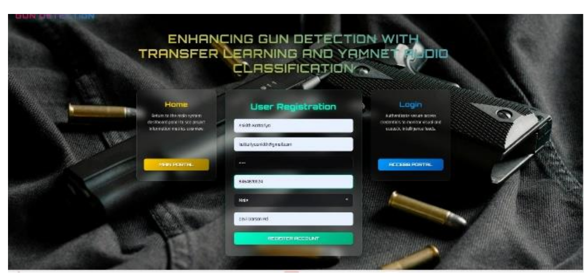


Fig. 3: Registration

Fig. 4: Dashboard

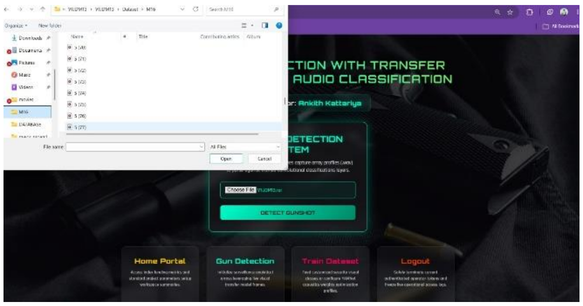


Fig. 5: Upload

Fig. 6: Predicted Results

V. CONCLUSION

The conclusion of our intelligent gun detection system demonstrates the successful implementation of a web-based firearm classification platform using Java Servlets, JSP, MySQL, MFCC feature extraction, and YAMNet-inspired audio analysis concepts. The project effectively identifies different gun types based on gunshot audio signals by extracting Mel Frequency Cepstral Coefficient features and comparing them with trained dataset values stored in the database. The developed system provides a secure authentication mechanism through user registration and login modules, ensuring authorized access to dataset training and gun detection functionalities. The bulk dataset training module efficiently processes multiple firearm audio samples and stores their acoustic feature representations for prediction purposes. The gun detection module successfully analyzes uploaded gunshot audio files and predicts the corresponding firearm category along with visual representation through the user interface. The interactive JSP-based dashboard and responsive UI design improve usability and provide an efficient environment for users to perform training and detection operations. The integration of audio processing techniques with machine learning concepts demonstrates the capability of intelligent systems in forensic, military, and defense applications. The system architecture ensures modularity, scalability, and efficient communication between presentation, processing, and database layers. The use of MFCC extraction and similarity-based prediction provides a lightweight and effective alternative for firearm audio classification using pure Java technology.

Experimental analysis with multiple gunshot audio samples highlights the effectiveness of the proposed approach in identifying firearm categories with considerable accuracy. The project also establishes a strong foundation for future integration of advanced deep learning models such as YAMNet, CNN, and LSTM for improved prediction performance. Overall, the developed system proves to be a reliable, intelligent, and scalable firearm audio detection platform capable of supporting modern security, surveillance, forensic investigation, and defense monitoring applications.

REFERENCES

[1] E. Tsalera, A. Papadakis, and M. Samarakou, ‘‘Comparison of pre-trained CNNsfor audio classification using transfer learning,’’ J. Sensor Actuator Netw., vol. 10, no. 4, p. 72, Dec. 2021.

[2] S. Patil and K. Wani, ‘‘Gear fault detection using noise analysis and machine learning algorithm with YAMNet pretrained network,’’ Mater. Today, Proc., vol. 72, pp. 1322–1327, 2023.

[3] W. Chen, H. Kamachi, A. Yokokubo, and G. Lopez, ‘‘Bone conduction eating activity detection based on YAMNet transfer learning and LSTM networks,’’ in Proc. 15th Int. Joint Conf. Biomed.Eng.Syst. Technol., 2022.

[4] S. Dogan, ‘‘A new fractal H-tree pattern based gun model identification method using gunshot audios,’’ Appl. Acoust., vol. 177, Jun. 2021, Art. no. 107916.

[5] R. Nijhawan, S. A. Ansari, S. Kumar, F. Alassery, and S. M. El-kenawy, ‘‘Gun identification from gunshot audios for secure public places using transformer learning,’’ Sci. Rep., vol. 12, no. 1, pp. 1–5, Aug. 2022.

[6] J. Park, ‘‘Enemy spotted: In-game gun sound dataset for gunshot classification and localization,’’ in Proc. IEEE Conf. Games (CoG), 2022, pp. 56–63.

AUTHORS Profile

Dr. M. V. Narayana received his Bachelor of Engineering (B.E.) degree in Computer Science in 2004 from JNTU Kakinada, India. He completed his Master of Technology (M.Tech) degree in Computer Science in 2010 from JNTU Hyderabad, India, and was awarded his Ph.D. in Computer Science in 2017 from JNTU Kakinada, India. With over 20 years of dedicated experience in academia and research, he currently serves as a Professor and Head of the Department (HOD) in the CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India. His research interests encompass various areas within computer science, and he has contributed significantly to the field throughout his career.

Mr. ANKITH KATARIA pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026..

Ms. U. JAHEERA BHANU pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.

Mr. P. SAI SRIKAR pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026

Mr. CH. HARI GANESH pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026