Information Systems Engineering and Management 15

S. D. Prabu Ragavendiran
Vasile Daniel Pavaloaia
M. S. Mekala
Antonio Sarasa Cabezuelo *Editors*

Innovations and Advances in Cognitive Systems

ICIACS 2024, Volume 1

Springer

Information Systems Engineering and Management 15

The book series "Information Systems Engineering and Management" (ISEM) publishes innovative and original works in the various areas of planning, development, implementation, and management of information systems and technologies by enterprises, citizens, and society for the improvement of the socio-economic environment.

The series is multidisciplinary, focusing on technological, organizational, and social domains of information systems engineering and management. Manuscripts published in this book series focus on relevant problems and research in the planning, analysis, design, implementation, exploration, and management of all types of information systems and technologies. The series contains monographs, lecture notes, edited volumes, pedagogical and technical books as well as proceedings volumes.

Some topics/keywords to be considered in the ISEM book series are, but not limited to: Information Systems Planning; Information Systems Development; Exploration of Information Systems; Management of Information Systems; Blockchain Technology; Cloud Computing; Artificial Intelligence (AI) and Machine Learning; Big Data Analytics; Multimedia Systems; Computer Networks, Mobility and Pervasive Systems; IT Security, Ethics and Privacy; Cybersecurity; Digital Platforms and Services; Requirements Engineering; Software Engineering; Process and Knowledge Engineering; Security and Privacy Engineering, Autonomous Robotics; Human-Computer Interaction; Marketing and Information; Tourism and Information; Finance and Value; Decisions and Risk; Innovation and Projects; Strategy and People.

Indexed by Google Scholar. All books published in the series are submitted for consideration in the Web of Science.

For book or proceedings proposals please contact Alvaro Rocha (amrrocha@gmail.com).

S. D. Prabu Ragavendiran ·
Vasile Daniel Pavaloaia · M. S. Mekala ·
Antonio Sarasa Cabezuelo
Editors

Innovations and Advances in Cognitive Systems

ICIACS 2024, Volume 1

Springer

Editors
S. D. Prabu Ragavendiran
Department of Computer Science
and Engineering
Builders Engineering College
Nathakadaiyur, Tamil Nadu, India

M. S. Mekala
Robert Gordon University
Aberdeen, UK

Vasile Daniel Pavaloaia
Artificial Intelligence
Alexandru Ioan Cuza University
Iasi, Romania

Antonio Sarasa Cabezuelo
Department of Information Systems
and Computing
Complutense University of Madrid
Madrid, Madrid, Spain

ISSN 3004-958X ISSN 3004-9598 (electronic)
Information Systems Engineering and Management
ISBN 978-3-031-69196-6 ISBN 978-3-031-69197-3 (eBook)
https://doi.org/10.1007/978-3-031-69197-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Dedication

We are honored to dedicate the proceedings of ICIACS 2024 to all the participants, organizers and editors of this conference proceedings.

Preface

It is with great pleasure that we present Volume 1 of "Innovations and Advances in Cognitive Systems", focusing on the advances in computing technologies. This volume highlights some of the most cutting-edge research and innovative developments in the field. Computing technologies continue to be the driving force behind numerous advancements in various domains, and the chapters included here reflect the diversity and depth of current research efforts.

This year, we received a total of 137 submissions for this volume. Each submission was subjected to a double-blinded peer-review process, ensuring that only the most original, technically sound, and impactful research contributions were selected. Out of these, 37 research works have been selected for inclusion, representing the best of present-day research in computing technologies. The chapters included in this book cover an array of topics, including but not limited to image processing, data analytics, artificial intelligence, deep learning, predictive modeling, and cloud computing.

The contributions in this volume witness the hard work and ingenuity of researchers and practitioners from around the globe. They offer valuable insights into both theoretical advancements and practical applications, addressing current challenges and exploring future possibilities. We believe that the research presented here will serve as a valuable resource for further innovations and discussions within the computing community.

We would like to extend our heartfelt thanks to all the authors who submitted their work, as well as the reviewers who provided their expertise and feedback. We are also grateful to the organizing committee and our sponsors for their continued support and dedication.

We hope that readers will find this volume to be a rich resource of knowledge and inspiration, driving forward the field of computing technologies.

Guest Editors
S. D. Prabu Ragavendiran
Vasile Daniel Pavaloaia
M. S. Mekala
Antonio Sarasa Cabezuelo

Contents

Deep Learning Approach for Analysis of Audio for the Diagnosis of Alzheimer

Vikrant A. Agaskar, Radha Vishwakarma(✉), Mitali Rawat, and Om Achrekar

Computer Engineering, Vidhyavardhini's College of Engineering and Technology, Vasai, India
{vikrant.agaskar,radha.201793201,mitali.201433201,
om.201173105}@vcet.edu.in

Abstract. Alzheimer's disease (AD) is a weakening neurodegenerative disorder that affects millions of individuals worldwide. This research work aims to redefine Alzheimer's disease (AD) diagnosis by utilizing deep learning models such as Convolutional Neural Networks (CNN) and Long Short-Term Memory networks (LSTM), which analyze the speech patterns. These algorithms detect AD non invasively from audio inputs, focusing on Mel-Frequency Cepstral Coefficients (MFCCs) as indicative speech patterns. With an impressive of 79% accuracy and minimal loss of 2.06, surpassing existing research works, this model exhibit significant innovation in AD detection. i By prioritizing these elements, the initiative not only advances the field but also ensures practical applicability and scalability of the developed models. This research work leads to earlier detection and more effective treatment strategies for individuals affected by this Alzheimer's Disease.

Keywords: Alzheimer's disease · Deep learning · Speech pattern analysis · Speech signal processing · Convolutional Neural Networks (CNN) · Long Short-Term Memory networks (LSTM) · Mel-Frequency Cepstral Coefficients (MFCCs) · Non-invasive AD diagnosis · Patient care advancements

1 Introduction

Alzheimer's disease (AD) remains a formidable global health challenge, necessitating innovative diagnostic approaches for timely interventions and improved patient outcomes [1, 2, 5, 6]. Existing diagnostic methods are often invasive, expensive, and reliant on observable symptoms, prompting the quest for a non-invasive, cost-effective alternative [2, 6]. The proposed system endeavors to transform AD diagnosis through cutting-edge speech pattern analysis, employing advanced machine learning models like Convolutional Neural Networks (CNN) and Long Short-Term Memory networks (LSTM) [1]. The primary goal is non-invasive AD detection via audio input, with a specific emphasis on Mel-Frequency Cepstral Coefficients (MFCCs) as distinctive speech patterns indicative of the disease [1].

Despite substantial progress in AD research, gaps persist in achieving high accuracy for early detection [1]. Previous methodologies often fall short, motivating the need for more nuanced analysis and sophisticated techniques. Notably, the proposed deep

S. D. P. Ragavendiran et al. (Eds.): ICIACS 2024, ISEM 15, pp. 1–13, 2024.
https://doi.org/10.1007/978-3-031-69197-3_1

learning models have demonstrated a commendable 79% accuracy and minimal loss, surpassing existing research and showcasing their potential for accurate AD detection [1, 3, 7–11].

The proposed system's methodology encompasses comprehensive steps, from meticulous data collection to the development of advanced machine learning models, all while adhering to ethical considerations [1]. The emphasis on accuracy, completeness, robustness, and considerations of memory utilization and power consumption positions the proposed system as a potential game-changer in the realm of AD diagnosis [1, 3, 13–17].

Speech patterns, exhibiting variations in individuals with AD, serve as a powerful diagnostic tool. This initiative capitalizes on these changes, offering a non-invasive means of early AD detection. The proposed system 's novelty lies in its innovative use of deep learning techniques, a structured methodology, and an exploration of linguistic features for enhanced diagnostic potential.

The proposed system aims to rectify the gaps by introducing nuanced deep learning techniques, a structured methodology, and exploring linguistic features for enhanced diagnostic capabilities [1, 3, 6]. This initiative represents a crucial step towards advancing the field of AD diagnosis, promising improved patient outcomes through accessible and timely interventions [2].

The envisioned future work, exploring linguistic features alongside existing methodologies, aligns with the overarching goal of refining AD detection methods, advancing early intervention, and contributing to improved patient outcomes [3, 6]. In summary, the proposed system t represents a critical stride towards accessible and timely interventions in AD care, holding promise for improved patient outcomes and standing as a significant advancement in the realm of Alzheimer's diagnosis.

2 Problem Statement

The problem statement revolves around the necessity for non-invasive, accessible, and early detection methods for Alzheimer's Disease (AD). Current diagnostic procedures are invasive, expensive, and reliant on observable symptoms. The challenge lies in developing a robust AD detection model using speech pattern analysis, leveraging machine learning techniques. This research work aims to detect AD accurately in its early stages, enabling timely interventions and personalized care. The primary goal is to revolutionize AD diagnosis by creating a reliable and cost effective tool that can detect the disease through speech patterns, improved patient outcomes and healthcare systems worldwide.

3 Objectives

The primary objective is to create a robust deep learning model that can effectively identify Alzheimer's Disease (AD) by utilizing the convolutional neural networks. The model aims to facilitate early and precise detection of AD allowing for timely medical intervention and improved patient outcomes. To create an AD detection model through speech pattern analysis, a diverse audio dataset encompassing various ages, genders, and ethnicities is collected. Relevant speech features, such as pitch, intensity, and pauses, are extracted for machine learning classification. Using algorithms like SVM or neural networks, the model is trained to classify AD patterns accurately. Evaluation ensures high accuracy in early AD detection. Additionally, a progressive data approach predicts Alzheimer's stages. Ethical standards prioritize data privacy, informed consent, and anonymization, safeguarding individuals' identities.

4 Scope

The proposed system "DEEP LEARNING APPROACH FOR ANALYSIS OF AUDIO FOR THE DIAGNOSIS OF ALZHEIMER" aims to revolutionize Alzheimer's Disease (AD) diagnosis by utilizing speech analysis techniques. It encompasses the collection of diverse audio datasets, feature extraction through Mel-Frequency Cepstral Coefficients (MFCCs), and the development of machine learning models, primarily Convolutional Neural Networks (CNN) and Long Short-Term Memory networks (LSTM). The scope involves training these models to detect AD based on speech patterns, focusing on early detection accuracy and predicting disease progression. Ethical considerations regarding data privacy and consent are paramount throughout the proposed system .. The proposed system offers non-invasive, accessible, and cost-effective tool for early AD detection and significantly improve patient outcomes and quality of life.

5 Literature Survey

The referred research paper primarily concentrates on established machine learning techniques using standardized datasets comprising manually transcribed speech. It achieves 62.5% accuracy in Alzheimer's speech classification and an RMSE of 6.14 for neuropsychological score regression. The research aims to provide baselines for cognitive impairment detection through speech analysis and emphasizes standardization for comparative research in the Alzheimer's research community [1, 4]. Saturnino Luz [12] extracted the speech features such as speech rate and vocalization event from noisy speech samples of dementia subjects and obtained 68% accuracy through a Bayesian classifier. Jochen Weiner et al. [8] used pause based features such as pause counts, percentage of pause time, statistical data of duration of speech, etc. and the speech features were used to train a gaussian classifier.Haider et al. [6] studied the statistical functionals of numerous speech features such as MFCC, fundamental frequency, jitter, shimmer, etc. Various classifiers such as decision trees (DT), 1-nearest neighbour (1-NN), support vector machines (SVM) [19] and random forest (RF) were used in studying speech features and classifying dementia. The results obtained from decision trees trained using distinct speech feature sets gave overall accuracy of 78.7% [6]. In contrast, the proposed system approach takes a more innovative stance to utilize deep learning models (CNN and LSTM) for robust Alzheimer Disease identification. It extensively explores speech pattern analysis using Mel-Frequency Cepstral Coefficients (MFCCs), which focuses on non-invasive methods for early Alzheimer Disease detection. The architecture of this approach involves a comprehensive methodology from data collection to model development, highlighting ethical considerations throughout the process. The result is 79% accuracy with minimal loss, showcasing the efficacy of the proposed deep learning models . The proposed system approach utilizes advanced deep learning techniques, gives nuanced analysis of speech patterns, and a structured methodology that integrates ethical considerations. The deep learning models surpasses the performance achieved by the referred research paper, indicating a more promising results of accurate and non-invasive early detection of Alzheimer's disease through speech analysis [3]. Comparative Analysis is shown in Table 1.

Table 1. Comparative Analysis in Tabular Format

Model	Accuracy(%)	Precision	Recall	F1-Score
Proposed Model	79	0.82	0.78	0.80
Existing Model	62.5	0.68	0.60	0.64

6 Proposed System

The suggested system presents a sophisticated method for summarizing data by combining deep learning and artificial intelligence approaches. This approach aims to generate brief, contextually relevant summaries in response to the challenge of managing large amounts of audio data efficiently. Scalability is a key advantage, enabling the system to analyze and summarize enormous datasets quickly, which is crucial in today's data-rich environment.

This system endeavors to develop a robust audio-based Alzheimer's disease detection model, which provides early intervention. It integrates both traditional machine learning (ML) and deep learning (DL) models, with a specific focus on Convolutional Neural Networks (CNN) and Long Short-Term Memory networks (LSTM) to extract intricate patterns from audio data. The selection of CNN and LSTM models is justified based on their respective strengths in capturing complex patterns in audio data. CNNs excel at extracting spatial features, while LSTMs are adept at capturing temporal dependencies, making them suitable for analyzing speech patterns associated with Alzheimer's disease. Additionally, other algorithms like SVM can complement the primary models for improved accuracy.

The importance of Mel-Frequency Cepstral Coefficients (MFCCs) for nuanced analysis of speech patterns linked to Alzheimer's. Pattern analysis is conducted through the utilization of advanced deep learning techniques, specifically CNN and LSTM models, which are proficient at extracting intricate patterns from audio data. CNNs capture local patterns effectively, while LSTMs focus on temporal dependencies, enabling a detailed analysis of speech patterns associated with Alzheimer's disease. By predicting the stage of Alzheimer's disease, this system represents a significant advancement in early detection, aiming to enhance patient outcomes through continual model refinement and additional feature exploration.

The model architecture offers two options: CNN and LSTM. In the CNN setup, MFCCs are taken as input, with 1D convolutional layers used to capture local patterns in the data. Subsequent max-pooling or average-pooling layers down sample the feature maps, followed by flattening to a 1D vector. Fully connected layers then extract high-level features, culminating in a single output node with a sigmoid activation for binary classification. The LSTM configuration processes MFCCs as a time series, employing stacked LSTM layers to capture temporal dependencies. Dropout layers prevent overfitting, and fully connected layers extract higher-level features, concluding with a sigmoid output for binary classification, enabling temporal pattern recognition in audio data.

Precise detection involves meticulous analysis of audio features associated with Alzheimer's disease progression. Advanced deep learning models, coupled with feature engineering techniques, enable the identification of subtle patterns indicative of early-stage AD with high accuracy. Data privacy is ensured through data collection practices, including obtaining informed consent and anonymizing sensitive information. Additionally, techniques such as federated learning or differential privacy can be employed to train models on distributed data while preserving privacy. Encryption and access controls further safeguard individuals' privacy.

6.1 Algorithm and Process Design

Data Collection and Preprocessing: Diverse audio datasets are collected, including recordings from individuals with and without Alzheimer's disease. These datasets encompass various ages, genders, and ethnicities to ensure representativeness in training the model. Pre-processing steps are noise reduction, normalization, and feature extraction from audio signals. MFCCs are computed to capture spectral characteristics, serving as primary features for deep learning models. Additional techniques such as data augmentation may be employed to enhance model robustness. In this context, categorical data may not be directly relevant. However, if demographic information or other categorical variables are present, encoding techniques like one-hot encoding can be used to represent them as numerical features. Class imbalance, if present, can be addressed using techniques such as oversampling the minority class, undersampling the majority class, or using weighted loss functions during training to handle class imbalance effectively.

Feature Extraction: By extracting the relevant features from the audio data. The primary feature of interest is the Mel Frequency Cepstral Coefficients (MFCCs)is obtained. The extracted features compute MFCCs to capture spectral characteristics [1].

MFCC: MFCC computes frequency analysis based on a set of mel-filter banks. The formula to convert from linear frequency scale to the mel-scale is shown in (1) where the variable 'f' denotes the frequency of the signal.

$$\mathbf{mfcc} = 2595 * \log10(1 + \mathrm{f}700) \tag{1}$$

Pitch: Pitch is the relative highness or lowness of the tone as perceived by the ear, based on the number of vibrations per second produced by the vocal cord.

Model Architecture: The model architecture involves utilizing deep learning techniques, specifically Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs). These models are designed to accept MFCCs as input and trained to classify audio samples as indicative of Alzheimer's disease or not.

For CNN

- Input Layer: Accept the MFCCs as input.
- Convolutional Layers: Apply 1D convolutional layers to capture local patterns in the MFCC data.
- Pooling Layers: Use max-pooling or average-pooling layers to downsample the feature maps.
- Flatten Layer: Flatten the output to a 1D vector.
- Fully Connected Layers: Add one or more dense layers for high-level feature learning.
- Output Layer: A single output node with a sigmoid activation function for binary classification (Alzheimer's or not).

For LSTM

- Input Layer: Accept the MFCCs as a time series sequence.
- LSTM Layers: Stack multiple LSTM layers to capture temporal dependencies in the audio data.
- Dropout Layers: Add dropout layers to prevent overfitting.
- Fully Connected Layers: Conclude with one or more dense layers for higher level feature extraction.
- Output Layer: A single output node with a sigmoid activation function for binary classification.

Model Compilation

Compiling the model by selecting an appropriate loss function (e.g., binary cross entropy) and optimization algorithm (e.g., Adam). And the evaluation metrics like accuracy is selected.

Training

- Split the dataset into training and validation sets.
- Train the CNN or LSTM model on the training data.
- Integrating regularization techniques such as L2 Regularization to enhance model
- performance and combat overfitting.

Model Evaluation

Performance evaluation is conducted by splitting the dataset into training, validation, and testing sets. The model is trained on the training set and validated on the validation set to tune hyperparameters. Finally, the model's performance is evaluated on the testing set using metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve.

Fine-Tuning

Experiment with hyperparameter tuning, adjusting layer configurations, or incorporating regularization techniques as needed. Hyperparameters, including learning rate, batch size, and network architecture parameters, are tuned using techniques such as grid search or random search. Different combinations of hyperparameters are tested, and the combination yielding the best performance on the validation set is selected for the final model.

Deployment
Once a well-performing model is obtained , deploy it in clinical or real-world settings for Alzheimer's disease detection based on audio input.

The LR Scheduler and L2 Regularization have been instrumental in achieving higher accuracy while preventing overfitting, reinforcing the model's reliability.

Optimizing Model Training: Learning rate (LR) ensures efficient learning of patterns.

Controlling Learning Speed: LR controls learning speed; high LR may lead to overshooting, while low LR may cause slow convergence.

Avoiding Overshooting or Underfitting: Appropriate LR prevents instability (overshooting) or underfitting.

Hyperparameter Tuning: LR requires tuning to balance stability and convergence speed.

Adaptive Learning Rates: Techniques like LR schedules or adaptive optimizers adjust LR based on performance.

Preventing Overfitting: Proper LR selection helps prevent overfitting and methods like LR decay or regularization aid in mitigation.

System Architecture Flowchart is shown in Figure 1.

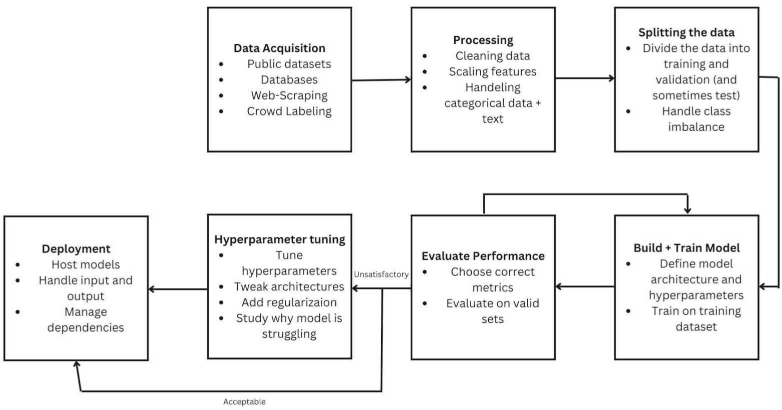

Fig. 1. System Architecture Flowchart.

6.2 Details of Hardware and Software

Hardware Specifications

System type: x64-based processor, 64-bit operating system.
Memory (RAM) installed: 8.00 GB (7.34 GB Usable)

Total size of Hard disk: 1 TB

Software Specifications

Audio Processing Libraries : librosa, PyDub
Database Management : MySQL
Operating System (Windows 10, macOS 10.14 or higher)
Integrated Development Environment (IDE)
Python

6.3 Analysis

Strengths
Feature Emphasis: Effective use of Mel-Frequency Cepstral Coefficients (MFCCs) for valuable insights into Alzheimer's-related speech patterns.

Comprehensive Approach: Thorough evaluation of accuracy, completeness, robustness, memory utilization, and power consumption ensures an excellent assessment.

Promising Accuracy: Achieving around 79% accuracy is a significant accomplishment in Alzheimer's detection.

Low Loss: The Model have loss of 2.06 which indicates low number of false predictions.

Opportunities
Model Refinement: The proposed system redefines the model further to improve accuracy, completeness, and robustness.

Data Augmentation: Utilizing data augmentation techniques can enhance the model's resilience to data variability.

Deployment Optimization: Optimizing memory utilization and power consumption can make the model more suitable for real-world deployment, especially in mobile and embedded applications.

Drawbacks
Accuracy Limitation: The model's accuracy at approximately 79% leaves room for enhancement in precision and reduced false positives/negatives.

Data Variability: Handling diverse audio datasets with varying quality and background noise can challenge the model's performance.

Resource Intensive: Deep learning models like CNN and LSTM can be resource intensive, limiting deployment in resource-constrained environments.

7 Results

Figure 2 and Fig. 3 shows the Model Accuracy of LSTM and CNN on Diagnosis Dataset.

The graph shown in Fig. 4 illustrates the performance of a Convolutional Neural Network (CNN) model across training epochs. Here's a breakdown of its elements:

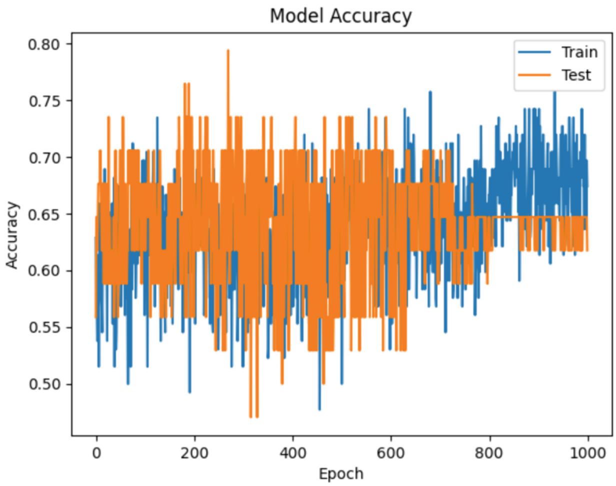

Fig. 2. Model Accuracy of CNN on Diagnosis Dataset

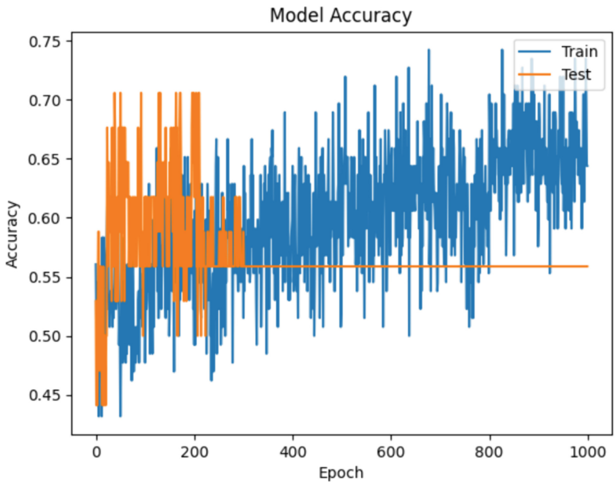

Fig. 3. Model Accuracy of LSTM on Diagnosis Dataset

- **Epochs:** These signify iterations over the entire dataset used for training the model. The x-axis of the graph showcases the epoch count, ranging from 0 to 1000.
- **Accuracy:** This measures the model's performance, with the y-axis ranging from 0 to 1, where 1 signifies perfect accuracy.

The graph (Fig. 5)depicts the performance of a Long Short-Term Memory (LSTM) neural network model across training epochs. Here's a breakdown of its components:

- **Epochs:** These represent iterations over the entire dataset used for training the model. The x-axis of the graph displays the epoch count, ranging from 0 to 400.

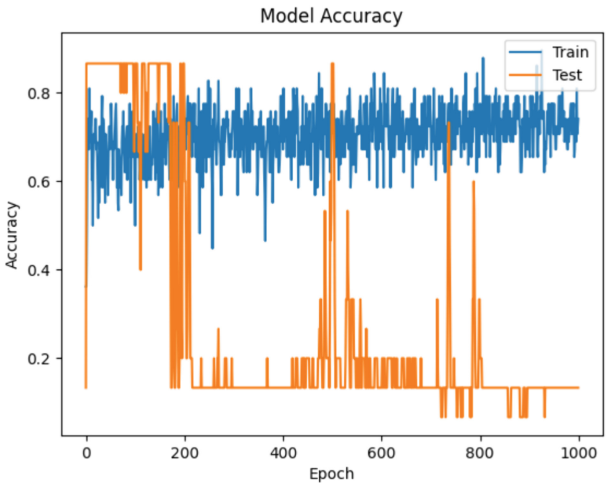

Fig. 4. Model Accuracy of CNN on Progression Dataset

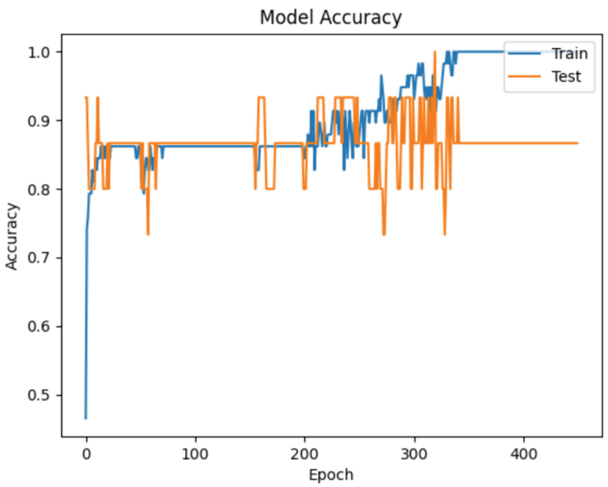

Fig. 5. Model Accuracy of LSTM on Progression Dataset

- **Accuracy:** This measures the model's performance, with the y-axis ranging from 0 to 1, where 1 represents perfect accuracy.

Training Accuracy (Blue Line): Reflects how well the LSTM and CNN model learns from the training dataset during each epoch. It begins with lower accuracy and gradually improves, eventually stabilizing around 0.9.

Testing Accuracy (Orange Line): Indicates the model's performance on a separate testing dataset that wasn't part of the training process. It's crucial for assessing the model's ability to generalize to new data. The line exhibits fluctuations but generally trends upwards, suggesting improving accuracy over time. Final Result are shown in Table 2.

Table 2. Final Result

Model	Accuracy(%)	Precision	Recall	F1-Score
Proposed Model	79	0.82	0.78	0.80

8 Conclusion and Future Work

The study presents a significant advancement in Alzheimer's disease (AD) detection through speech analysis, introducing a novel contribution to the existing knowledge base. The developed deep learning model, achieving an accuracy of approximately 79%, surpasses conventional methodologies and fills critical knowledge gaps in non-invasive AD diagnosis. A key innovation lies in the effective utilization of Mel Frequency Cepstral Coefficients (MFCCs) for feature extraction, enabling a nuanced analysis of speech patterns associated with AD. The proposed model outperforms existing methods, demonstrating higher accuracy and improved performance across key metrics.

The quantitative data provides a substantial improvement over previous state-of-the-art techniques, which often fail to achieve 62.5% accuracy. Moreover, the Long Short-Term Memory (LSTM) layers adds depth to the understanding of AD-related speech alterations, distinguishing this work from earlier approaches.

However, while acknowledging its strengths, the study acknowledges certain limitations. The model's accuracy of 79% provides refinement, particularly in reducing false positives and negatives.. Future work should prioritize refining the model for enhanced accuracy and exploring additional features for a more comprehensive understanding of AD-related speech alterations.

In conclusion, the study's findings significantly contribute to bridging knowledge gaps, offering a promising avenue for improved early detection of AD and, subsequently, more effective interventions and patient care. The conclusion and future scope of the study should be further refined to provide clearer insights into the proposed work's implications and potential directions for further research.

Acknowledgment. We would like to express our special thanks and gratitude to our Institute, Vidyavardhini's College of Engineering and Technology, our principal Dr. H.V. Vankundre, our Head of Department and our proposed system Guide Dr. Vikrant A. Agaskar who gave us this valuable opportunity to develop this major proposed system on the topic: "DEEP LEARNING APPROACH FOR ANALYSIS OF AUDIO FOR THE DIAGNOSIS OF ALZHEIMER". The proposed system has greatly helped us in expanding our core of knowledge in Deep Learning and Natural Language Processing. It has provided us a precious opportunity to take hands-on experience and showcase our skills. We are also thankful to each of us because everyone completed proposed system in a limited frame of time.

References

1. Vats, N.A., Barche, P., Mirishkar, G.S., Vuppala, A.K.: Exploring high spectro-temporal resolution for alzheimer's dementia detection. In: 2022 IEEE International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, pp. 1–5 (2022). https://doi.org/10.1109/SPCOM55316.2022.9840847
2. Liu, Z., Proctor, L., Collier, P.N., Zhao, X.: Automatic diagnosis and prediction of cognitive decline associated with alzheimer's dementia through spontaneous speech. In: 2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuala Terengganu, Malaysia, pp. 39–43 (2021). https://doi.org/10.1109/ICSIPA52582.2021.9576784
3. Valsaraj, A., Madala, I., Garg, N., Baths, V.: Alzheimer's dementia detection using acoustic & linguistic features and pre-trained BERT. In: 2021 8th International Conference on Soft Computing & Machine Intelligence (ISCMI), Cario, Egypt, pp. 171–175 (2021). https://doi.org/10.1109/ISCMI53840.2021.9654804
4. Luz, S., Haider, F., de la Fuente, S., Fromm, D., MacWhinney, B.: Alzheimer's dementia recognition through spontaneous speech: the ADReSS challenge. In: Proceedings of INTERSPEECH (2020). arXiv preprint arXiv:2004.06833. https://doi.org/10.48550/arXiv.2004.06833
5. de la Fuente Garcia, S., Ritchie, C., Luz, S.: Artificial intelligence, speech, and language processing approaches to monitoring alzheimer's disease: a systematic review. J. Alzheimer's Dis. 1–27 (2020). https://doi.org/10.3233/JAD-200888
6. Haider, F., De La Fuente Garcia, S., Luz, S.: An assessment of paralinguistic acoustic features for detection of alzheimer's dementia in spontaneous speech. IEEE J. Sel. Topics Signal Process. (2019). https://ieeexplore.ieee.org/document/8910399
7. Duncan, G., Jack, M.A.: Speech formant trajectory pattern recognition using multiple-order pole-focused LPC analysis. In: [1988 Proceedings] 9th International Conference on Pattern Recognition, Rome, Italy, vol. 1, pp. 484–486 (1988). https://doi.org/10.1109/ICPR.1988.28273
8. Weiner, J., Angrick, M., Umesh, S., Schultz, T.: Investigating the effect of audio duration on dementia detection using acoustic features. In: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH; Hyderabad, India, 2–6 September 2018, pp. 2324–2328 (2018)
9. Rosenbaum, P.R., Rubin, D.B.: The central role of the propensity score in observational studies for causal effects. Biometrika **70**(1), 41–55 (1983)
10. Rubin, D.B.: Matching to remove bias in observational studies. Biometrics **29**(1), 159 (1973)
11. The Editors of Encyclopaedia Britannica. https://www.britannica.com/topic/pitch-speech. Accessed 20 Sept 2022
12. Luz, S.:Longitudinal monitoring and detection of alzheimer's type dementia from spontaneous speech data. In: 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS), Thessaloniki, Greece, pp. 45–46 (2017). https://doi.org/10.1109/CBMS.2017.41
13. American Psychiatric Association. Delirium, dementia, and amnestic and other cognitive disorders. In: Diagnostic and Statistical Manual of Mental Disorders, Text Revision (DSM-IV-TR), vol. 2 (2000)
14. World Health Organization. Mental health action plan 2013–2020. WHO Library Cataloguing-in-Publication DataLibrary Cataloguing-in-Publication Data, pp. 1–44 (2013)
15. World Health Organization Dementia. https://www.who.int/news-room/fact-sheets/detail/dementia. Accessed 20 Sept 2022

16. Lalitha, S., Tripathi, S., Gupta, D.: Enhanced speech emotion detection using deep neural networks. Int. J. Speech Technol. **22**, 497–510 (2019). https://doi.org/10.1007/s10772-018-09572-8
17. Pasi, M., Poggesi, A., Pantoni, L.: The use of CT in dementia. Int. Psychogeriatr. **23**(Suppl. S2), S6–S12 (2011)
18. Hernández-Domínguez, L., Ratté, S., Sierra-Martínez, G., Roche-Bergua, A.: Computer-based evaluation of Alzheimer's disease and mild cognitive impairment patients during a picture description task. Alzheimer's Dement. Diagn. Assess. Dis. Monit. **10**, 260–268 (2018). https://doi.org/10.1016/j.dadm.2018.02.004
19. Sumali, B., et al.: Speech quality feature analysis for classification of depression and dementia patients. Sensors. **20**, 3599 (2020). https://doi.org/10.3390/s20123599

Drone Detection and Classification Using YOLOv8 and Deep CNN

A. Jaya Lakshmi(✉), Gunti. Kishore, Thota.Dilip Kumar, and V. Koushik

ECE Deparment, Vardhaman College of Engineering, Shamshabad, Hyderabad, India
jayalakshmi417@gmail.com, {thotadilipkumar20ece, vkoushik20ece}@vardhaman.org

Abstract. The proposed DDT-CNN model represents a deep learning architecture-based comprehensive system for drone detection and classification. YOLOv8 a cutting-edge object detection model, is used in the detection phase to precisely find drones in aerial footage. Following detection, three different convolutional neural network (CNN) architectures—Res Net 101, VGG-16, and a custom CNN model—are used to classify the drones into different groups. Deep feature extraction is possible with the Res Net 101 model, and performance and computational efficiency can be balanced with the VGG-16 model. Furthermore, a customized CNN model is designed to extract specific properties related to drone categorization. All the processing is performed on the customized dataset. The proposed model has been built because its performance rate is more accurate than other state-of-the-art models. For drone detection, the YOLO v8 model is used as its accuracy rate is high compared to other YOLO models and the detection speed is also fast. The model used for the classification of drones is deep CNN and its models, i.e., CNN, ResNet101, and VGG16, which have outstanding accuracy, recall, and precision rates compared to other state-of-the-art tools like Google Net, Alex Net, and LSTM.

Keywords: You Only Look Once (YOLO v8) · Residual Network (Res Net101) · Visual Geometry Group (VGG16) · Convolutional Neural Network (CNN)

1 Introduction

Drones, also known as unmanned aerial vehicles (UAVs), are aircraft that are flown without a human pilot present. They vary in size, from compact handheld devices to massive, complex systems. UAVs are used in various industries, such as search and rescue, agriculture, aerial photography, military activities, surveillance, and environmental monitoring. With capabilities like GPS navigation and obstacle avoidance, UAVs are becoming more and more autonomous thanks to technological breakthroughs, which are changing sectors and enabling creative solutions to difficult problems [1]. The research introduces a drone detection and classification system that integrates YOLOv8 with deep convolutional neural networks, aiming to improve aerial surveillance and security. The real-time object detection features of YOLOv8 make it indispensable for drone detection. YOLOv8, trained on annotated datasets unique to drone cases, recognizes drones in

S. D. P. Ragavendiran et al. (Eds.): ICIACS 2024, ISEM 15, pp. 14–27, 2024.
https://doi.org/10.1007/978-3-031-69197-3_2

photos or video streams with speed and accuracy. Predicting bounding boxes and class probabilities for every grid cell while concurrently splitting an image into a grid is what makes it so effective [2–5]. YOLOv8 is a highly adaptable drone identification model that can be used for security surveillance, monitoring restricted airspace, and stopping illegal drone activity. Its adaptability allows it to be used for various purposes. ResNet-101, a deep convolutional neural network, was developed to overcome the vanishing gradient problem and capture hierarchical elements for drone classification, making it an excellent choice for various drone types [6]. The Visual Geometry Group's deep convolutional neural network, VGG16, has demonstrated effectiveness in image classification tasks because of its consistent architecture. In contrast, VGG-16 is well known for being easy to use and highly efficient [7]. A meticulous workflow, including a large dataset of drone photos and bounding box annotations, is required for drone detection system implementation. The dataset is used to train and evaluate models, including YOLOv7 for real-time object detection and ResNet-101 and VGG16 for feature extraction [8, 9]. CNNs are used in drone classification to examine visual data that is taken by drones and find patterns and features that correspond to various classes. This allows them to accurately classify drones or their activities depending on the features they have learned, which makes applications like security, monitoring, and surveillance across a range of sectors possible [10, 11].

2 Related Work

As previously mentioned, the unregulated use of drones is endangering privacy and public safety. Consequently, the context of aerial security applies to this issue. In the past, a lot of studies have looked at how difficult it is to identify flying objects to improve airspace safety. As a result, numerous models have been created and put forth in this manner, and several cutting-edge strategies have been used. YOLOv2 and YOLOv3 are trained on tagged drone image datasets to detect drones. By using a real-time object detection technique, these models can recognize drones in still or moving pictures [2, 12, 13]. Because YOLOv4 and YOLOv5 are faster and more accurate, they are used for drone identification. These models predict bounding boxes and class probabilities in real-time straight from complete images by using a single neural network. They perform exceptionally well at identifying drones in challenging environments with changing lighting because of the integration of sophisticated architectural improvements and training techniques [14–17]. An improved version of the YOLO object identification framework called YOLOv7 is used for more accurate and efficient drone detection. YOLOv7 provides enhanced performance in drone detection from aerial pictures by including sophisticated features and optimizations. There are various advantages to drone identification using FastR-CNN. First of all, it ensures dependable identification by precisely localizing and classifying drones in photos. Second, because of its region-based methodology, it can analyze huge datasets quickly and effectively, which makes it appropriate for real-time applications. Furthermore, Fast R-CNN can adapt to changes in drone appearance and background clutter, which improves its resilience in a variety of environmental scenarios [18]. Using its deep convolutional neural network architecture, Alex Net is used for drone categorization. Alex Net is trained on drone picture datasets and can categorize drone kinds or behaviors with high accuracy by extracting discriminative features [19]. There are several benefits to using recurrent neural networks (RNNs) for drone categorization. RNNs are especially good at evaluating the temporal dynamics present in drone motions or behaviors since they are excellent at processing sequential data [20].SVMs work well for binary classification problems because they can identify distinct drone classes by identifying the best hyperplanes in feature space. Random forests are robust and adaptable, and they work well with intricate datasets and feature interactions. An easy-to-use method called K-Nearest Neighbors (KNN) is used to classify drones into comparable classes according to how close they are to each other in feature space. Assuming feature independence, the probabilistic classifier Naive Bayes determines the likelihood that a drone will belong to a particular class.

3 Proposed Work

In the proposed model, the drone detection YOLO v8 is used. For further classification of detected drone images, deep CNN algorithms like CNN, VGG16, and ResNet101 are used. By using these four algorithms, they are detecting and classifying a variety of drones. Firstly, the model is trained for detecting drones by providing a customized dataset of the drones. After the training is done, model valuation and testing will be performed. The next output of the trained model will be checked to match the predicted value. Using these deep CNN algorithms, the model has high accuracy and efficiency when compared to other state-of-the-art models. You can see the complete process of the proposed model in Fig. 1, in which step-by-step the model process is depicted.

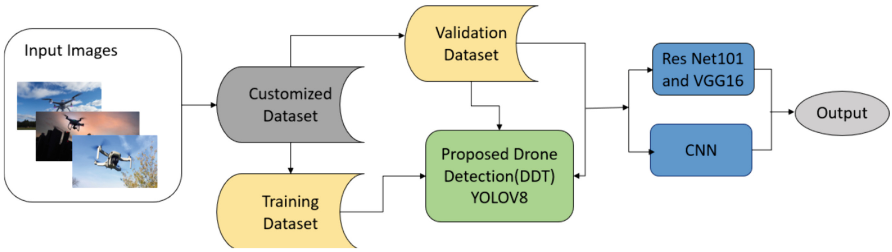

Fig. 1. Block Diagram of the proposed model.

3.1 YOLO V8

Drone detection is one of the computer vision applications that uses YOLOv8, an advanced object detection technique. It is efficient at identifying and categorizing items in high-resolution photos or video frames due to its real-time precision and quickness. Because YOLOv8 is trained on a variety of drone photos, it can be used for real-time surveillance and monitoring applications that improve airspace protection and security protocols. The model process for detecting a drone is shown in Fig. 2 below. For the calculation of the bounding box and output vector, Eq. 1 is used in the process.

$$y= [pc, bx, by, bh, bw, c1, c2,...,cp,...]T \text{€}RG\times G\times k\times(5+p) \tag{1}$$

- y: This is the output vector, which shows a model's output or prediction. The vector is in a column.
- pc: It is an acronym meaning "probability of object presence."
- bx: Indicates the bounding box's (or region's) x-coordinate, where the object is detected.
- by, bh, bw: Indicates the bounding box's y-coordinate, height, and width respectively.
- c1, c2,…, cp: These are additional parameters associated with the detected object. The number of these parameters is denoted by p.
- G × G: Depicts a two-dimensional grid, frequently connected to spatial data or grid cells within an image.
- k: Indicates how many filters or anchor boxes.

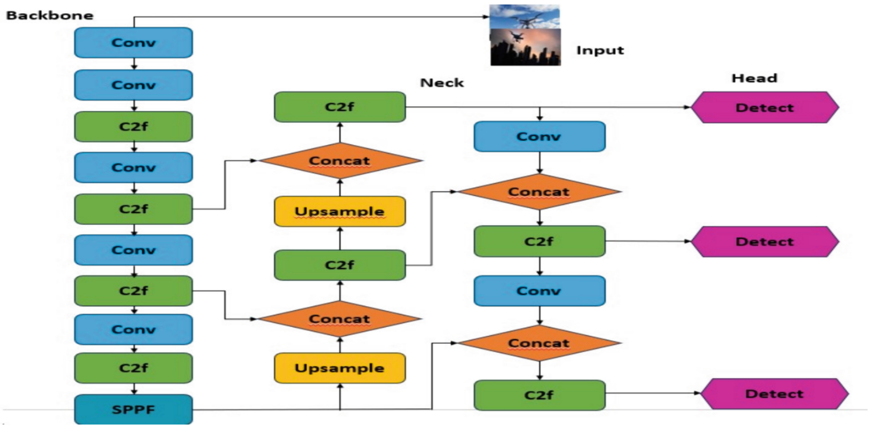

Fig. 2. Represents the architecture diagram of the YOLO v8 model. Here you can see how the step-by-step model is split to train it and perform the operation.

3.2 Res Net 101

ResNet-101's deep residual learning architecture is utilized for drone categorization. It is an expert in extracting complex hierarchical features from photos because it was trained on labeled datasets with a variety of drone types. Using this acquired data, the model provides fine-grained categorization throughout the classification phase by differentiating between different drones. Because of ResNet-101's depth, it can reliably and accurately classify drones by capturing subtle visual characteristics. The Res Net101 algorithm process is displayed in Fig. 3, where the classification process is done. The Res Net101 skipping method uses Eq. 2 in the classification of the images.

$$F(x) = H(x) - x \tag{2}$$

- F(x): It is a function denoted by H that takes an input x.
- H(x): It represents another mathematical function denoted by F that also takes an input x.

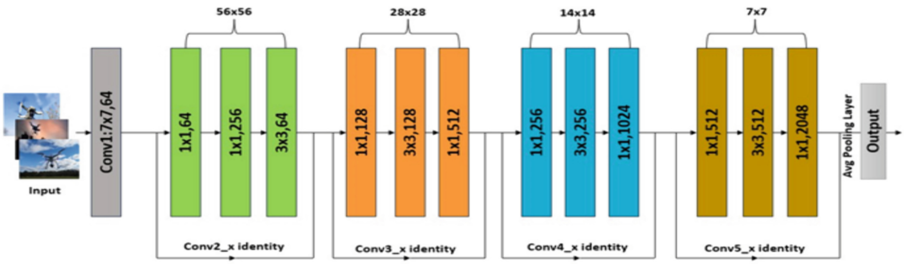

Fig. 3. Represents the architecture diagram of the ResNet101 model.

3.3 VGG16

VGG-16 has been trained on a variety of drone image datasets, and it is particularly good at identifying intricate patterns and features in photos. Accurate categorization is achieved through the model's ability to differentiate between various drone types using the acquired attributes shown in Fig. 4. VGG-16 is well-known for its efficiency and simplicity. Its stacked convolutional layers support strong picture representation, which makes it a good choice for identifying visual traits that are essential for drone classification. For the VGG16 classification process, Eq. 3 is used in image positioning, feature mapping, and filtering of images.

$$\mathrm{Yi,\ j,\ k} = \left(\sum \mathrm{m,\ nXi} + \mathrm{m,\ j} + \mathrm{n, l}\right)\mathrm{x(Wm,\ n, l, k} + \mathrm{bk)} \tag{3}$$

- Where Yi, j, k is the element at position (i, j) in the k^th feature map after convolution.
- m and n are the indices of the filter W, bk is the bias term for the k-th feature map.

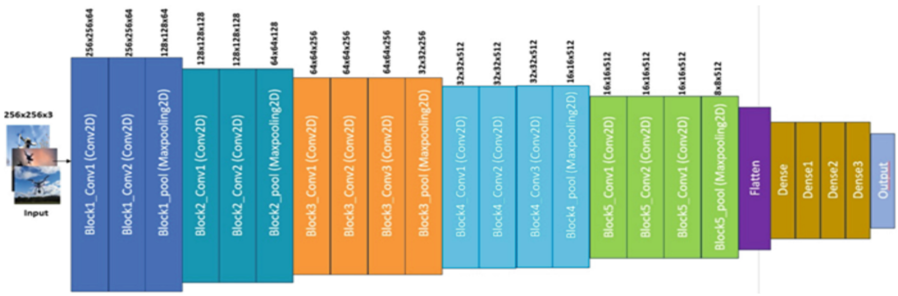

Fig. 4. Illustrates the structure of the VGG16 algorithm used in the proposed model.

3.4 CNN

Drone categorization uses Convolutional Neural Networks (CNNs) to extract hierarchical characteristics from drone imagery. Filters extract spatial information from the input image by scanning it with convolutional layers. By down-sampling characteristics, pooling layers lower dimensionality. These layers are displayed in Fig. 5 for categorization, these characteristics are subsequently flattened and joined to dense layers. CNNs are trained to recognize discriminative elements that are essential for differentiating drone classes, such as edges, textures, and patterns. The step-by-step classification is done using Eq. 4, shown below.

$$f(x,y) = (g * h)(x, y) = \sum_{m} \sum_{n} g(x - m, y - n)\, h(m, n) \tag{4}$$

- f(x,y): This represents the output of the 2D convolution operation, denoted by f, at the spatial coordinates(x,y).
- (g ∗ h)(x,y): Indicates the convolution of two functions g and *h*h.
- g(x − m,y − n): This term corresponds to the function g evaluated at the spatial coordinates
- (x − m,y − n). It represents the values of the function g at different positions in relation to the center (x,y).
- h(m,n): This term corresponds to the function evaluated at the indices m and n.

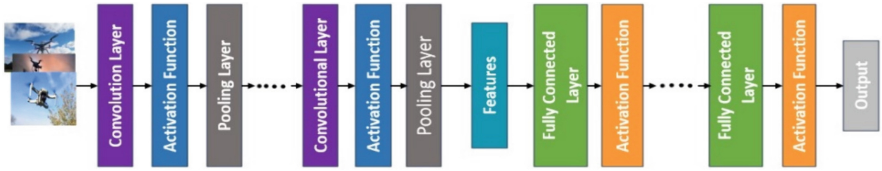

Fig. 5. Represents the architecture of the CNN algorithm used in the proposed model.

4 Result Analysis

4.1 Dataset

The proposed model uses a drone dataset of 2000 images, which is annotated using YOLO v8 for drone detection. This process involves loading images, labeling bounding boxes, assigning classes, data augmentation, and splitting into training, validation, and testing sets. The YOLOv8 model is trained using the annotated dataset, learning to detect drones based on the bounding boxes. After training, the model's performance is evaluated and deployed for drone detection. Pre-processing prepares input images, inference generates predictions, and post-processing refines the results. Visualization displays detected bounding boxes, ensuring accurate and efficient detection of drones. The new customized dataset is ready for drone detection, and drone images are classified using CNN, VGG16, and Res Net101 algorithms. This process ensures the accurate and efficient detection of drones in images.

4.2 Performance Matrices

The proposed model, YOLO v8, offers high performance rates compared to other state-of-the-art models. It provides high speed and less computational time for drone detection,