IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, June 2024

GAN-Enhanced Vocal Biomarker Analysis for Respiratory Health Assessment

Prof. Shweta Bhelonde, Abhinav Pandey, M. Rahul Surya, Onkar Bante, Divya Dongare, Mangesh Yadav, Anshul Rahate
Department of Artificial Intelligence Engineering
G H Raisoni Institute of Engineering and Technology, Nagpur, Maharashtra, India

Abstract: *Nearly two centuries ago, people became aware that various diseases, such as the common cold, asthma, Alzheimer's, and psychological disorders, manifest changes in a human voice. The recent emergence of the virus known as "COVID-19" has claimed millions of lives due to delayed detection of infected individuals. Traditional medical techniques for virus detection are time-consuming and costly. However, recent advancements in Artificial Intelligence (AI) offer remote diagnosis for analysing and identifying diseases that cause variations in voice. The evolution of machine learning provides numerous techniques to extract meaningful information from vocal biomarkers.*
This study explores innovative techniques to enhance the analysis of vocal biomarkers, emphasizing Generative Adversarial Networks (GANs) and machine learning for assessing respiratory diseases. The end goal of the study is to improve the performance by utilizing synthetic data for training purposes. Subsequently, machine learning models are employed to analyze real-time data for detecting respiratory illnesses. Comparing different machine learning algorithms gives us a better understanding of their capabilities and drawbacks

Keywords: Generative Adversarial Network (GAN), Wasserstein GAN, Conditional GAN, Artificial Intelligence (AI), Cough Detection, Respiratory Health Assessment, Vocal Biomarkers, MFCC, Mel-Spectrogram, Croma, Machine Learning, L2-regularization, Classification, Normalization, Support Vector Machine (SVM), Covid-19, Convolutional Neural Network (CNN), Audio Synthesis, Synthetic Data Generation, LSTM, Synthetic Minority Over-Sampling Technique (SMOTE), Deep Learning, Recurrent Neural Network (RNN), Data Augmentation, Cross-Validation, Principle Component Analysis (PCA), Zero Crossing Rate (ZCR)

I. INTRODUCTION

Voice analysis has long been recognized as a valuable tool in detecting various diseases, ranging from the common cold to psychological disorders, as alterations in voice can serve as early indicators of underlying health conditions. With the recent global outbreak of the COVID-19 virus, the need for efficient and timely detection methods has become more critical than ever. Traditional medical techniques for virus detection have proven to be both time-consuming and expensive, leading to significant delays in identifying infected individuals and preventing the spread of the virus. However, recent advancements in Artificial Intelligence (AI) and machine learning have opened up new possibilities for remote diagnosis, offering a promising avenue for analyzing and identifying diseases based on variations in voice patterns.

In this context, this study aims to explore innovative techniques for enhancing the analysis of vocal biomarkers, emphasizing leveraging Generative Adversarial Networks (GANs) and machine learning algorithms for assessing respiratory diseases. By harnessing the power of synthetic data for training purposes, the study seeks to improve the performance and accuracy of disease detection models. By applying machine learning models to real-time data, the study aims to create reliable diagnostic tools capable of flawlessly analyzing respiratory illnesses.

By comparing and evaluating different machine learning algorithms, the study aims to gain insights into their respective strengths and limitations, ultimately contributing to a better understanding of their abilities for disease detection. The

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, June 2024

findings of this study have the potential to revolutionize the field of medical diagnostics, offering more efficient and reliable methods for early disease detectionand intervention.

II. METHODOLOGY

2.1 Dataset Collection

While gathering data for this study, a meticulous approach of assembling a comprehensive dataset comprising audio recordings of individuals afflicted with various respiratory ailments. These utilized recordings were from reputable online repositories, including those established by organizations such as the United Nations. A spectrum of respiratory conditions such as asthma, diabetes, cough, cold, smoking-related illnesses, and especially COVID-19 status, curating the dataset to provide a holistic representation of respiratory health challenges.

The COSWARA dataset is a large crowd-sourced collection of cough, breathing, and speech recordings, along with participant metadata, aimed at developing computational tools for COVID-19 detection and monitoring using vocal biomarkers, while addressing challenges related to data diversity and quality.

Special attention was directed towards ensuring diversity within the dataset, with deliberate efforts to incorporate individuals spanning different age groups, genders, and ethnic backgrounds. By embracing diversity, the dataset aims to mirror the real-world heterogeneity observed among individuals affected by respiratory diseases. This inclusivity serves as a foundational element in bolstering the robustness and applicability of the models developed in this study. Due to the association of diverse demographic profiles, our models are poised to yield more nuanced and accurate insights, leisurely enhancing their relevance and effectiveness in practical settings. This commitment to comprehensive dataset collection underscores our dedication to advancing research in respiratoryhealth diagnostics through ethically sound and inclusive data acquisition practices.

2.2 Dataset Cleaning and Preprocessing

Undergoing data preprocessing involved a systematic and rigorous approach to ensuring the quality and suitability of the collected audio dataset for subsequent analysis. Initially, in the cleaning phase, a meticulous approach of identifying and eliminating any corrupted files or instances of inconsistent dimensionality, whichever could probably disrupt model performance. This thoroughcleaning process constituted a crucial precursor to subsequent preprocessing steps, laying the foundation for robust and reliable data analysis. Moreover, to accommodate the variability in the lengthof audio files within the dataset, careful consideration was given to maintaining a flexible input length strategy. At moments when neural networks necessitated fixed-length inputs for processing, padding was engaged judiciously to standardize input dimensions and facilitate seamless data processing by adding or removing elements to ensure a consistent length. Additionally, shifting techniques were employed to augment the data and introduce variations, enhancing the model's ability to generalize. Moreover, proactive measures were implemented to address potential anomalies, such as datasets containing zero-length files, through the implementation of exception-handling protocols. Followingthe cleaning, length standardization, and data augmentation processes, normalization techniques such as Peak Normalization, RMS (Root Mean Square) Normalization, Min-Max Normalization and Z-scorenormalization (Standardization) were used on the audio files to ensure uniformity and consistency in data representation. This normalization step played a pivotal role in enhancing the comparability and interpretability of the dataset across different samples.

2.3 Generative Adversarial Network (GAN):

A tailored GAN architecture was meticulously crafted to generate synthetic audio data, enriching our dataset with diverse and realistic samples representative of respiratory disease recordings.Trained on meticulously cleaned and pre-processed datasets, this GAN model underwent rigorous validation to ensure that the generated synthetic data closely resembled real respiratory recordings, thereby enhancing the dataset's variability and authenticity. Comprising a generator and discriminator, the GAN architecture employed sequential layers including dense, convolutional, and flatten layers within the generator to generate synthetic files, while the discriminator discerned between synthetic and original files. To introduce non-linearity and promote better learning, the LeakyReLU activation function was adeptly utilized, leveraging its characteristic small slope for negative values to prevent stagnation and encourage feature extraction. Additionally, the Adam optimizer was employed to dynamically adjust the learning rate for each parameter,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, June 2024

particularly beneficial in navigating complex loss landscapes and stabilizing the training process to facilitate convergence toward optimal solutions. Throughout the training process, the dataset was meticulously fed into the GAN model in predefined batch sizes during each epoch, with discriminator and generator losses calculated iteratively to gauge model performance and refinement. This comprehensive approach to data preprocessing and GAN model training ensured the generation of synthetic data that closely mirrored real-world respiratory recordings, thereby enriching the dataset with diverse and authentic samples for subsequent analysisand model training.

2.4 The Standard Generative Adversarial Network (SGAN)

The Standard Generative Adversarial Network (SGAN) employs adversarial training to generate synthetic data that mirrors the distribution of the original dataset. Structurally, SGANs feature two neural networks: a Generator and a Discriminator. Operating within a zero-sum game framework, these networks engage in competition, where one's improvement results in the other's detriment. The Generator's objective is to fabricate realistic data instances capable of deceiving the Discriminator. By processing random noise vectors, it creates artificial samples, striving to maximize the likelihood of the Discriminator classifying them as authentic. Conversely, the Discriminator aims to differentiate between real and fake data, gauging the probability of samples being genuine or synthetic. Through adversarial training, both networks refine their performance iteratively, adjusting parameters based on their respective loss functions. Ideally, as training progresses, the Generator excels at crafting lifelike samples, while the Discriminator becomes adept at discerning authenticity. Ultimately, at convergence, the Generator produces outputs indistinguishable from real data, while the Discriminator struggles to reliably distinguish between genuine and fabricated samples. SGAN had limitation on the audio dataset and could not produce results as expected. Therefore, we moved our research to various advancementin the GAN architectures.

2.5 Wasserstein Generative Adversarial Network (WGAN)

In contrast to SGAN, the Wasserstein Generative Adversarial Network (WGAN) introduces the Earth Mover's distance (Wasserstein distance) to enhance training stability and alleviate mode collapse. Its architecture resembles SGANs, comprising a Generator and a Discriminator network. However, the focus shifts to minimizing the Wasserstein distance between the distributions of real and generated samples. The Generator endeavors to craft synthetic data approximating the underlying dataset's distribution, employing random noise vectors as input. Through training, it aims to diminish the Wasserstein distance, ensuring generated samples closely resemble real data. Conversely, the Discriminator's role involves estimating the Wasserstein distance, discerning between authentic and synthetic samples effectively. Adversarial training persists, with both networks striving to improve their performance iteratively, driven by the Wasserstein distance and gradients. Over time, the Generator refines its ability to generate samples closely resembling real data, while the Discriminator enhances its accuracy in estimating the Wasserstein distance. At convergence, the Generator produces outputs resembling authentic data, and the Discriminator accurately estimates the Wasserstein distance betweenreal and synthetic samples.

2.6 Conditional Generative Adversarial Networks (CGANs)

Conditional Generative Adversarial Networks (CGANs) extend the GAN framework by incorporating conditional information to guide the generation process. Like SGANs and WGANs, CGANs feature a Generator and a Discriminator network. Additionally, they intake conditional information, such as class labels, alongside random noise vectors. The Generator's objective remains unchanged: to produce realistic samples conditioned on the provided information. It leverages both random noise vectors and conditional information to generate synthetic data samples, aiming for in distinguishability from real data instances. The Discriminator, informed by the conditional information, discerns between real and fake data, classifying samples accordingly. Adversarial training ensues, where both networks vie for improved performance. The Generator endeavours to generate samples faithful to the conditional distribution, while the Discriminator enhances its ability to discriminate between genuine and synthetic samples. As training progresses, the Generator becomes adept at producing realistic samples aligned with the provided conditioning, while the Discriminator struggles to reliably distinguish between real and synthetic samples given the conditioning. At

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, June 2024

convergence, the Generator successfully matches the conditional distribution, while the Discriminator's ability to differentiate between real and synthetic samples diminishes.

2.7 Feature Extraction/Engineering:

Feature extraction is like simplifying audio by picking out the most important parts. In this discussion, we'll explore why feature extraction matters and how different methods help us do it effectively. Understanding these methods helps us improve audio technology for various applications.

- **Mel-frequency cepstral coefficients (MFCCs)** are a cornerstone in the field of speech and audio processing, revered for their ability to capture the nuanced power spectrum of sound over short durations. Modelled after the human auditory system's response to audio frequencies, MFCCs serve as a robust feature set for tasks ranging from speech recognition to speaker identification. Their effectiveness lies in their capability to extract complex audio signals into a compact representation, making them indispensable in various real-world applications.
- **Mel-spectrograms** offer a visual depiction of the power spectrum of audio signals, with a twist that aligns more closely with human auditory perception. By converting frequencies to the mel scale, this representation provides a nuanced insight into pitch perception. Widely employed in tasks such as music genre classification and sound event detection, mel-spectrograms empower researchers and practitioners to delve into the intricate patterns of audio data, facilitating more accurate analyses and interpretations.
- **Chroma** features illuminate the distribution of energy across different musical notes or pitch classes, offering a unique lens into the harmonic content of audio signals. Renowned for their robustness in handling variations in timbre and instrumentation, chroma features find applications in diverse domains such as music genre classification, chord recognition, and melody extraction. Their ability to capture the essence of musical content makes them invaluable tools for researchers and musicians alike.
- **Zero Crossing Rate (ZCR)** stands as a simple yet powerful metric, revealing insights into the frequency content and periodicity of audio signals. By quantifying the rate at which an audio waveform crosses the zero-amplitude axis, ZCR provides valuable information for tasks like speech recognition, music genre classification, and audio fingerprinting. Its versatility and efficiency make it astaple feature extraction method in the arsenal of audio signal processing techniques.
- **Root Mean Square Error (RMSE)** serves as a fundamental measure for assessing the fidelity of reconstructed audio signals compared to their originals. Widely employed in audio compression algorithms and quality assessment tasks, RMSE quantifies the average magnitude of differences between predicted and actual values. Its application extends beyond signal processing, offering valuable insights into the perceptual quality of audio data and guiding improvements in audio processing algorithms and technologies.

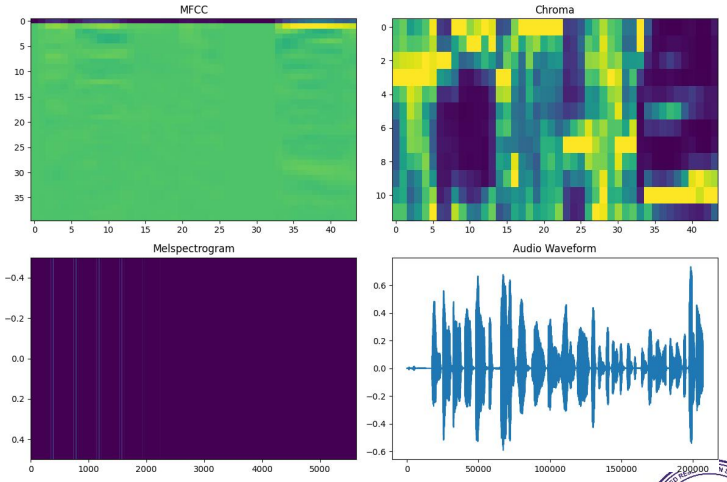

Fig. - Various features present in the audio dataset.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, June 2024

The image provides a detailed analysis of an audio signal through four visualizations. The top- left plot shows Mel-Frequency Cepstral Coefficients (MFCCs) with color intensity indicating magnitude changes over time. The top-right plot illustrates Chroma features, depicting energy distribution across 12 pitch classes. The bottom-left plot is a Mel-spectrogram, showing amplitude variations over the Mel frequency scale, though it appears mostly dark, indicating minimal signal presence. The bottom-right plot displays the raw audio waveform, highlighting amplitude variations over time. Together, these plots offer a comprehensive view of the audio signal's characteristics.

2.8 Machine learning & Deep Learning algorithm for classification:

In the pursuit of achieving optimal performance for a specific task, a comprehensive exploration of various machine learning and deep learning algorithms was undertaken. Initial experimentation with machine learning algorithms revealed challenges stemming from the high dimensionality of the data, hindering their ability to effectively capture and utilize intricate patterns present within the dataset. Despite efforts to fine-tune these algorithms, they struggled to yield promising results, prompting a strategic shift towards leveraging the capabilities of deep learning models.

With a focus on deep learning algorithms, which inherently possess greater capacity for learning complex representations from high-dimensional data, the investigation delved into their potential to address the challenges encountered with machine learning approaches. This transition allowed for a more nuanced exploration of the dataset's intricacies, empowering the models to extract meaningful features and representations that could better inform decision-making processes.

Furthermore, to augment the available training data and potentially enhance model generalization, both real-world data and synthetic data generated through a Generative Adversarial Network (GAN) were incorporated into the training pipeline. This approach sought to capitalize on the synthetic data's ability to introduce diversity and variability into the dataset, thereby potentially mitigating issues related to overfitting and improving the model's ability to generalize to unseen data instances.

Throughout the iterative process of model development, rigorous optimization techniques and hyperparameter tuning played a pivotal role in fine-tuning the models' parameters. The primary objective was twofold: to maximize classification accuracy on the training data and to foster robust generalization capabilities to ensure the model's efficacy in real-world scenarios. By meticulously optimizing model parameters and tuning hyperparameters, the aim was to strike a balance between model complexity and generalization performance, ultimately yielding models that could effectively navigate the complexities inherent in high-dimensional data spaces.

In essence, the exploration of machine learning and deep learning algorithms, coupled with the integration of real and synthetic data, underscored the importance of a holistic approach to model development. Through this approach, insights were gained into the strengths and limitations of different algorithmic strategies, paving the way for more informed decision-making processes and the advancement of model performance in the face of high-dimensional data challenges.

Here are the brief of the Machine Learning and Deep Learning models used in this project:

2.9 Machine Learning Model

Machine Learning Models are computational algorithms designed to identify patterns and make decisions based on data. These models can be broadly classified into three categories: supervised learning, where models are trained on labeled data to predict outcomes; unsupervised learning, which identifies hidden patterns in unlabeled data; and reinforcement learning, where models learn to make decisions through trial and error by receiving rewards or penalties. Common types of machine learning models include linear regression, decision trees, support vector machines, neural networks, and clustering algorithms like k-means. These models are essential in various applications such as image and speech recognition, natural language processing, predictive analytics, and autonomous systems, driving significant advancements in technology and industry. There are various machine learning models utilized such as:

Support Vector Machine (SVM): SVM is a supervised learning model used for classification and regression tasks. It finds the hyperplane that best separates different classes in the feature space. SVM is effective for small to medium-sized datasets with high-dimensional feature spaces.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, June 2024

Random Forest: Random Forest is an ensemble learning method consisting of multiple decision trees. It constructs multiple trees during training and outputs the class that is the mode of the classes or mean prediction of the individual trees. Random Forest is known for its robustness and ability to handle large datasets with high dimensionality.

Principal Component Analysis (PCA): Principal Component Analysis (PCA) is an unsupervised machine learning technique widely used for dimensionality reduction. It transforms a large set of correlated variables into a smaller set of uncorrelated variables called principal components, effectively summarizing the essential information of the dataset. PCA begins by standardizing the data and computing the covariance matrix to capture feature relationships. By calculating the eigenvalues and eigenvectors of this matrix, PCA identifies the directions (principal components) that capture the most variance in the data. The original data is then projected onto this new feature space, resulting in a reduced number of dimensions that retain most of the original dataset's variability. This process simplifies complex datasets, reduces computational costs, and helps in visualizing high-dimensional data while mitigating overfitting, making PCA an invaluable tool in data preprocessing and analysis.

DEEP LEARNING MODEL

Deep learning models are neural networks with multiple hidden layers. They learn complex patterns and representations directly from raw data, requiring large amounts of data and computational resources for training. Deep learning models often achieve state-of-the-art performance in various classification tasks.

To enhance the performance and generalization of these models, several techniques and architectures are utilized:

- **L2 Regularization:** This technique helps prevent overfitting by adding a penalty to the loss function based on the magnitude of the model's weights. It encourages the model to keep the weights small, promoting simpler models that generalize better to new data.
- **Synthetic Minority Over-sampling Technique (SMOTE):** SMOTE is used to address class imbalance by generating synthetic examples of the minority class. This helps the model learn better representations of the minority class, improving classification performance in imbalanced datasets.
- **Long Short-Term Memory (LSTM):** LSTMs are a type of recurrent neural network (RNN) that are particularly effective for sequential data. They can capture long-term dependencies and are widely used in time-series forecasting, natural language processing, and other tasks involving sequential data.
- **Training Without Up-sampling:** In some cases, models are trained without up-sampling the minority class, relying instead on the model's ability to learn from the available data. This can be complemented with techniques like class weighting or focal loss to handle class imbalance.
- **Fixed Class Imbalance:** When dealing with fixed class imbalance, techniques such as class weighting, where higher weights are assigned to minority classes, or using specialized loss functions like focal loss, can help the model focus more on the underrepresented classes.
- **Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN):** CNNs are highly effective for image and spatial data due to their ability to capture local patterns through convolutional layers. RNNs, including LSTMs, are ideal for sequential and temporal data, capturing dependencies over time.
- **Data Augmentation with Cross-Validation:** Data augmentation techniques, such as rotating, flipping, and scaling images, help increase the diversity of the training dataset, improving the model's robustness. Using 5-fold cross-validation ensures that the model is trained and validated on different subsets of the data, providing a more reliable estimate of its performance.

By incorporating these advanced techniques and architectures, deep learning models can be tailored to specific tasks and datasets, achieving superior performance and generalization.

Transformer Classifier:

The Transformer architecture revolutionized natural language processing tasks with self-attention mechanisms. It captures long-range dependencies in sequential data efficiently. Transformer models excel in various classification tasks, especially in natural language understanding and generation.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, June 2024

III. EVALUATION AND RESULT ANALYSIS

In the evaluation of the trained classification models for respiratory disease classification, a rigorous examination was conducted utilizing standard evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics served as fundamental tools in assessing the models' performance across diverse disease categories, ensuring a comprehensive understanding of their classification capabilities. By scrutinizing these metrics, we gained insights into the models' ability to correctly classify instances and discern potential areas for improvement. Moreover, the use of standard metrics allowed for a direct comparison of model performance, enabling us to identify the most effective approaches in disease classification.

A significant aspect of our evaluation strategy involved a comparative analysis to assess the impact of using GAN-generated synthetic data versus authentic, real-world data for model training. This analysis aimed to elucidate any disparities in performance stemming from the utilization of synthetic data augmentation techniques. By juxtaposing the performance of models trained on synthetic data against those trained on real data, we aimed to uncover the potential benefits and limitations of synthetic data augmentation in the context of medical classification tasks. Through this comparative analysis, we aimed to provide valuable insights into the efficacy of synthetic data in improving model performance and generalization across different disease categories.

The results of our evaluation were meticulously analyzed to identify trends, patterns, and insights into the performance and efficacy of various machine learning algorithms in respiratory disease classification. By delving deep into the observed outcomes and scrutinizing the underlying methodologies, we gained a nuanced understanding of the strengths and weaknesses of different algorithms. This in-depth analysis not only provided valuable insights into the performance of individual algorithms but also facilitated informed decision-making regarding the selection and optimization of algorithms for future iterations of the classification system. Overall, our evaluation framework enabled us to gain valuable insights into the performance of classification models for respiratory disease classification, paving the way for potential enhancements and advancements in the field.

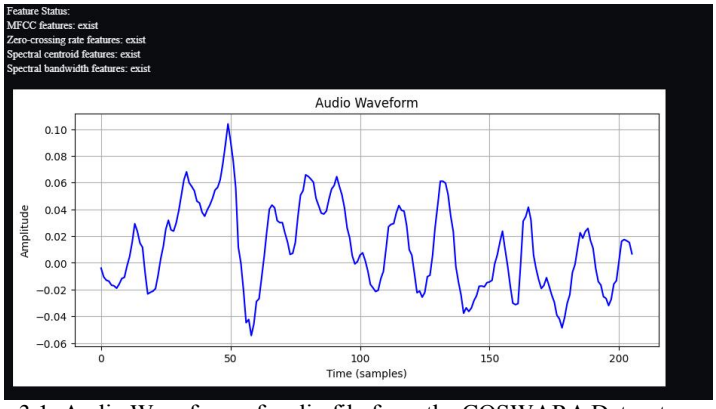

Fig.3.1- Audio Waveform of audio file from the COSWARA Dataset.

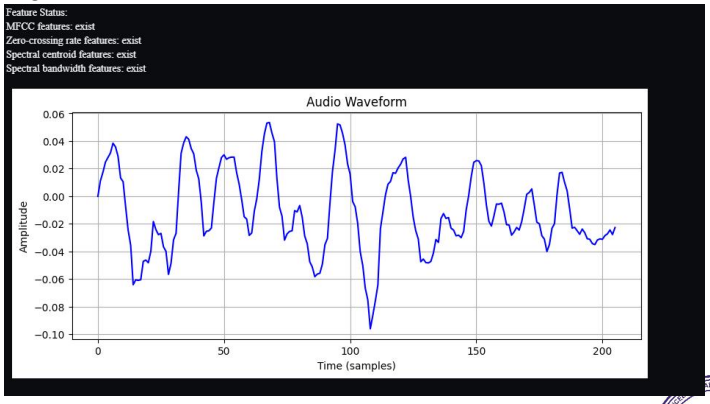

Fig.3.2- Audio Waveform of GAN Generated audio Dataset.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, June 2024

Both figures display audio waveforms, which are graphical representations of the amplitude (vertical axis) of sound waves over time (horizontal axis). The top figure labelled "Fig 3.1 - Audio Waveform of audio file from the COSWARA Dataset," shows a waveform with distinct peaks and troughs, indicating variations in the amplitude of the audio signal. The bottom figure labelled "Fig 3.2 - Audio Waveform of GAN Generated audio Dataset," exhibits a waveform with a slightly different pattern, suggesting variations in the characteristics of the audio signal compared to the first dataset. These waveform visualizations can be useful for analysing and understanding the properties of audio data.

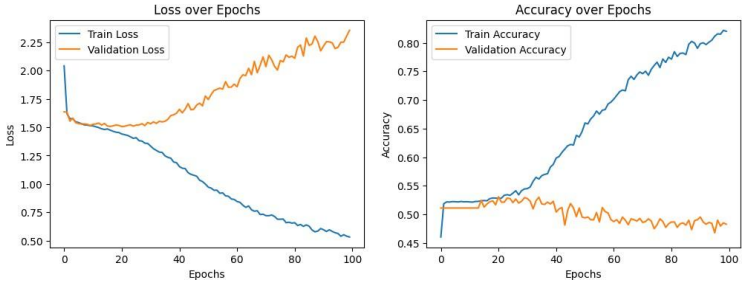

Fig.3.3- Accuracy of Deep Learning Classification Model Without using SMOTE.

Figure 3.3 contains two line graphs that depict the performance of a deep learning classification model without using SMOTE (Synthetic Minority Over-sampling Technique). The left graph shows theloss over epochs, with the training loss (orange line) and validation loss (blue line) plotted against the number of epochs. The right graph shows the accuracy over epochs, with the training accuracy (orange line) and validation accuracy (blue line) plotted against the number of epochs. Both graphs suggest thatthe model's performance improves as the number of epochs increases, with the loss decreasing and accuracy increasing

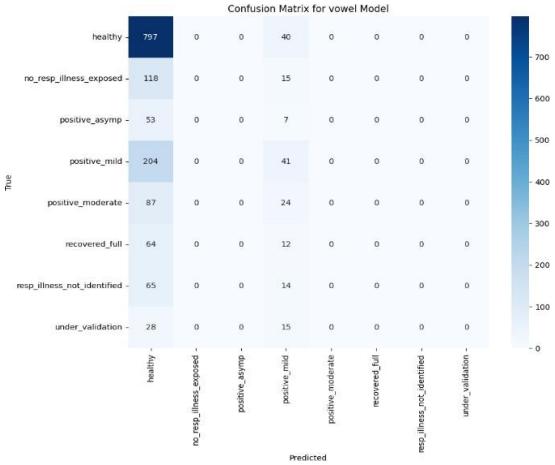

Fig.3.4- Confusion Matrix of Deep Learning Classification Model Without using SMOTE.

Figure 3.4 presents a confusion matrix for the deep learning classification model without using SMOTE. The confusion matrix is a table that displays the actual and predicted classes for the model's predictions. The rows represent the true classes, while the columns represent the predicted classes. The diagonal elements (shaded in blue) represent the correctly classified instances, while the off-diagonal elements indicate misclassifications. The confusion matrix provides insights into the model's performance, highlighting which classes are being misclassified and to what extent, allowing for further analysis and potential improvements to the model.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, June 2024

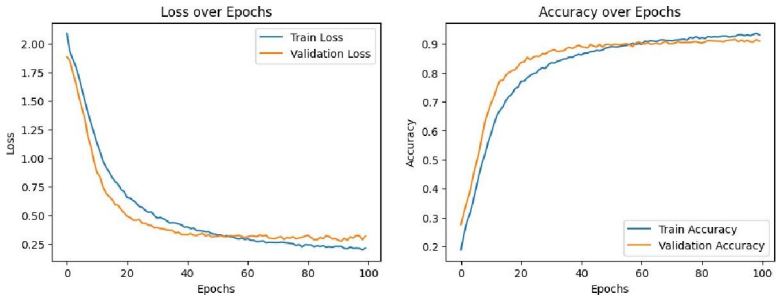

Fig.3.5- Accuracy of Deep Learning Classification Model using SMOTE.

Figure 3.5 shows two line graphs that illustrate the performance of a deep learning classificationmodel using SMOTE (Synthetic Minority Over-sampling Technique). The left graph displays the loss over epochs, with the training loss (orange line) and validation loss (blue line) plotted against the number of epochs. The right graph depicts the accuracy over epochs, with the training accuracy (orangeline) and validation accuracy (blue line) plotted against the number of epochs. Both graphs indicate thatthe model's performance improves as the number of epochs increases, with the loss decreasing and accuracy increasing, similar to the previous model without using SMOTE

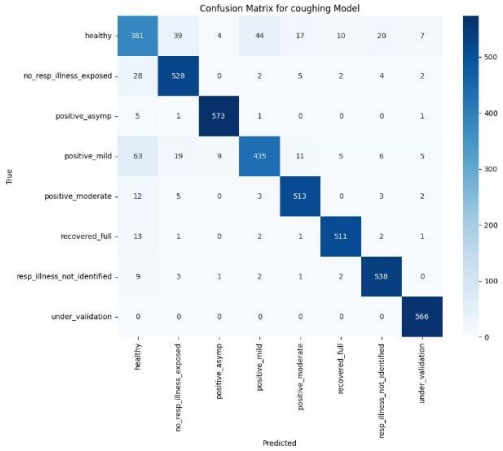

Fig.3.6- Confusion Matrix of Deep Learning Classification Model using SMOTE.

Figure 3.6 presents a confusion matrix for the deep learning classification model using SMOTE.The confusion matrix is a table that shows the actual and predicted classes for the model's predictions. The rows represent the true classes, while the columns represent the predicted classes.

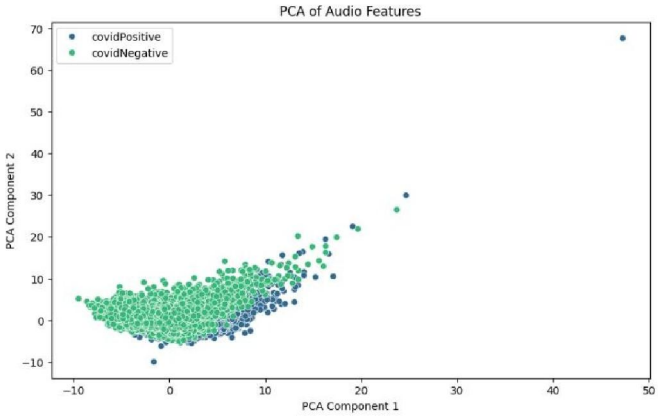

Fig.3.7- Classification using PCA (Principle Component Analysis).

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, June 2024

The diagonal elements (shaded in blue) represent the correctly classified instances, while the off-diagonal elements indicate misclassifications. By comparing this confusion matrix with the previous one without using SMOTE, insights can be gained into the impact of SMOTE on the model's performance and the changes in misclassification patterns across different classes.

In Figure 3.7, PCA visualizes clusters but shows overlap between green dots and blue triangles, indicating its limitation in separating classes. This underscores the need for advanced methods like GANs for generating synthetic data to enhance class separability, and deep learning models that handle nonlinear relationships directly from high-dimensional features. These approaches offer more robust classification capabilities compared to PCA's linear dimensionality reduction

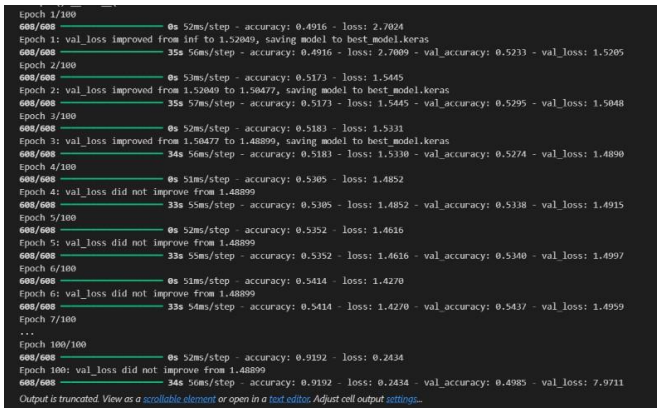

Fig.3.8 - Classification Model Accuracy.

Figure 3.8 presents a list of classification model accuracies, likely corresponding to different machine learning models or configurations. Each row displays the model's name, accuracy value, and potentially other metrics or parameters. The accuracy values range from around 0.51 to 0.97, indicating varying levels of performance across the different models or configurations. This type of tabular representation allows for easy comparison and evaluation of multiple classification models based on their achieved accuracies or other reported metrics.

Classification Model Evaluation Table

Model Name	Sound Types	Evaluation Metrics
Deep Learning	Vowels	52%
Deep Learning using L2 Regularization	ALL	91.92%
Deep Learning using SMOTE	Coughing	91%
	Breathing	89.56%
	Vowels	87.94%
	Counting	89.6%
Deep Learning using LSTM	Coughing	98.6%
SVM with RBF Kernel	ALL	53.4%
SVM with Sigmoid Kernel	ALL	43.84%
LDA (Linear Discriminant Analysis)	ALL	51.63%
Deep Learning without Up-sampling	Coughing	Training- 96.33% Testing- 85.32%
Deep Learning with Fixed Class Imbalance	Coughing	Training- 99.9%
		Testing- 99.4%
Deep Learning Network Model (MFCC, CFCFT, Mel Spectrogram)	Coughing	96.73%
CNN & RNN(LSTM)	Coughing	96.77%
		Fold 1- 97.82%

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, June 2024

		Fold 2- 96.82% Fold 3-97.82% Fold 4- 59.01% Fold 5- 94.53%
Data Augmented Deep Learning Model with 5 Cross-Folds	Coughing	Cross-Val Accuracy-89.19%

Table 3.1- Accuracy of different Classification Models on various sound types.

IV. DISCUSSION AND FUTURE DIRECTION

Discussion

The study's findings have significant implications for healthcare diagnostics, particularly in the realm of respiratory disease detection. By integrating Generative Adversarial Networks (GANs) and machine learning techniques in vocal biomarker analysis, the potential benefits of leveraging advanced computational methods in medical diagnostics are highlighted. Specifically, the incorporation of GANs enables the generation of synthetic data that exhibits higher accuracy with anomalous, huge, and diverse datasets. This augmentation of the dataset not only enhances the robustness of machine learningmodels but also improves their performance in detecting respiratory diseases.

However, due to limited computational resources, the use of GANs had to be dropped, and the focus was shifted to solidifying the classification using various machine learning models. The utilization of GANs in conjunction with machine learning algorithms offers promising avenues for enhancing the accuracy and reliability of diagnostic tools in healthcare. By harnessing the power of synthetic data generated by GANs, healthcare professionals can access a more comprehensive and diverse dataset, leading to more accurate disease detection and diagnosis. Moreover, the integration of vocal biomarker analysis holds potential for non-invasive and cost-effective disease screening, revolutionizing the field of respiratory disease diagnostics.

The classification model evaluation table presents the evaluation metrics for different classification models applied to various sound types. Deep learning architectures, including those incorporating techniques like L2 regularization, SMOTE (Synthetic Minority Over-sampling Technique), and LSTM (Long Short-Term Memory), achieved high accuracy scores ranging from 87.94% to 91.92%. Additionally, the CNN & RNN(LSTM) model and the Deep Learning Network Model (MFCC, CFCFT, Mel Spectrogram) demonstrated impressive performance with accuracies of 96.77% and96.75%, respectively, on coughing sounds

While GAN-based data augmentation could not be implemented due to computational constraints, alternative strategies were employed to enhance the robustness and reliability of the classification models. Cross-validation and fixed class imbalance handling were utilized as alternatives to GAN- based data augmentation.

Overall, this research showcases the potential of machine learning techniques in vocal biomarker analysis for respiratory disease diagnostics, even without the incorporation of GANs. The achieved classification accuracies across various sound types and models highlight the promising avenues for developing accurate and cost-effective diagnostic tools in the healthcare domain.

FUTURE ANALYSIS

The findings of this study pave the way for several promising avenues for future research. One potential direction involves exploring additional features that can further enhance the performance of machine learning models in respiratory disease detection. By incorporating novel biomarkers or physiological indicators, researchers can aim to refine the diagnostic capabilities of existing models and improve their accuracy in identifying respiratory conditions.

Additionally, future research efforts could focus on refining GAN architectures to better suit the needs of medical data generation. This may involve optimizing GAN parameters, exploring novel architectures, or incorporating domain-specific knowledge to enhance the quality and diversity of synthetic data generated by GANs, thereby improving the robustness and generalization of machine learning models trained on this synthetic data.

Furthermore, expanding the dataset represents a crucial aspect of future research endeavors. By incorporating a larger and more diverse dataset encompassing a wide range of respiratory conditions and demographic factors, researchers can strive to improve model generalization and robustness. This expansion of the dataset will enable machine learning

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, June 2024

models to capture a broader spectrum of disease manifestations, ultimately leading to more accurate and reliable diagnostic tools for respiratory disease detection.

Exploring these avenues will contribute to the development of advanced computational methods for medical diagnostics, leveraging the potential of machine learning techniques and data augmentation strategies to revolutionize the field of respiratory disease detection and diagnosis.

V. CONCLUSION

The methodology adopted in this study presents a systematic and innovative approach for harnessing artificial intelligence techniques to elevate the analysis of vocal biomarkers in respiratory disease diagnosis. By integrating advanced computational methods such as machine learning and Generative Adversarial Networks (GANs), our research establishes a robust framework for extracting valuable insights from vocal data and enhancing diagnostic accuracy. This methodology not only offers a novel avenue for non-invasive disease detection but also provides healthcare practitioners with a powerful tool to expedite diagnosis and intervention processes.

The models developed and insights gleaned from this research carry substantial promise for revolutionizing healthcare diagnostics and ultimately improving patient outcomes. Through the integration of machine learning algorithms and GAN-generated synthetic data, our study demonstrates the potential to achieve higher accuracy and reliability in respiratory disease diagnosis. By leveraging the wealth of information embedded within vocal biomarkers, healthcare professionals can make more informed clinical decisions, leading to earlier detection of respiratory conditions and timely intervention strategies. These advancements have the potential to significantly impact patient care, facilitating proactive management of respiratory diseases and ultimately enhancing quality of life.

In conclusion, the findings of this study underscore the transformative potential of artificial intelligence in healthcare diagnostics, particularly in the realm of respiratory disease diagnosis. By leveraging advanced computational methods and insights derived from vocal biomarker analysis, our research lays the groundwork for a new era of precision medicine. Moving forward, further research and collaboration are warranted to refine and validate the proposed methodologies, ultimately translating them into real-world clinical applications. With continued innovation and investment in AI- driven healthcare solutions, we can envision a future where early disease detection and personalized treatment strategies become the cornerstone of modern healthcare practices, leading to improved patient outcomes and enhanced overall well-being.

REFERENCES

[1]. Li Li, Alimu Ayiguli, Qiyun Luan, Boyi Yang, Yilamujiang Subinuer, Hui Gong, Abudureherman Zulipikaer, Xuemei Zhong, “Prediction and Diagnosis of Respiratory Disease by Combining Convolutional Neural Network and Bi-directional Long Short- Term Memory Methods”, Digital Public Health 2022, MAY 2022, Vol 10, https://doi.org/10.3389/fpubh.2022.881234.

[2]. Panagiotis Kapetanidis, Fotios Kalioras, Constantinos Tsakonas, Pantelis Tzamalis, George Kontogiannis, Theodora Karamanidou, hanos G. Stavropoulos, and Sotiris Nikoletseas, “Respiratory Diseases Diagnosis Using Audio Analysis and Artificial Intelligence: A Systematic Review”, Sensors 2024, FEB 2024, 24(4), 1173; https://doi.org/10.3390/s24041173.

[3]. Alper Idrisoglu, MSc; Ana Luiza Dallora, PhD; Peter Anderberg, phd; Johan Sanmartin Berglund, MD, PhD, “Applied Machine Learning Techniques to Diagnose Voice-Affecting Conditions and Disorders: Systematic Literature Review”, JMIR Publications, 2023, Vol 25, 10.2196/46105.

[4]. Teruhisa Watase, BSc, MPH; Yasuhiro Omiya, ME, PhD; Shinichi Tokuno, MD, PhD, “Severity Classification Using Dynamic Time Warping– Based Voice Biomarkers for Patients With COVID-19: Feasibility Cross-Sectional Study”, JMIR Publications, 2023, Vol 8, 10.2196/50924.

[5]. Jessica Robin, John E. Harrison, Liam D. Kaufman, Frank Rudzicz, William Simpson, Maria Yancheva, “Evaluation of Speech Based Digital Biomarkers: Review and Recommendations”, OCT 2020, 10.1159/000510820.

[6]. Troncoso, Á.; Ortega, J.A.; Seepold, R.; Madrid, N.M. “Non-invasive devices for respiratory sound monitoring”. Procedia Comput. 2021, Vol 192 (3040-3048), https://doi.org/10.1016/j.procs.2021.09.076.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, June 2024

[7]. Ijaz, A.; Nabeel, M.; Masood, U.; Mahmood, T.; Hashmi, M.S.; Posokhova, I.; Rizwan, A.; Imran, A. "Towards using cough for respiratory disease diagnosis by leveraging Artificial Intelligence: A survey". Inform. Med. Unlocked, 2022, Vol 29, https://doi.org/10.1016/j.imu.2021.100832.

[8]. Kim, H.; Jeon, J.; Han, Y.J.; Joo, Y.; Lee, J.; Lee, S.; Im, S. "Convolutional neural network classifies pathological voice change in laryngeal cancer with high accuracy". J. Clin. Med. 2020, Vol 9, https://doi.org/10.3390/jcm9113415.

[9]. Payten C, Chiapello G, Weir K, Madill C. "Terminology and frameworks used for the classification of voice disorders: a scoping review protocol". JBI Evid Synth. 2021; Vol 19:454-462, 10.11124/JBIES-20-00066. 10.10.1016/j.cmi.2019.09.009.

[10]. Xi Y, Tian C L, Qian L. "A study of deep learning methods for de-identification of clinical notes in cross-institute settings". BMC Med Inform Decis Mak. 2019, Vol 5:232, 10.1186/s12911-019-0935-4.

[11]. Ian J. Goodfellow; Jean Pouget-Abadie; Mehdi Mirza; Bing Xu; David Warde-Farley; Sherjil Ozair; Aaron Courville; Yoshua Bengio. "Generative Adversarial Networks", arxiv, 2014, https://doi.org/10.48550/arXiv.1406.2661 .

[12]. Shujian Liao, Hao Ni, Marc Sabate-Vidales, Lukasz Szpruch, Magnus Wiese, Baoren Xiao. "Sig-Wasserstein GANs for Conditional Time Series Generation", arxiv, 2023, https://arxiv.org/pdf/2006.05421.

[13]. Pirmin Philipp Ebner, Amr Eltelt. "Audio inpainting with generative adversarial network", arxiv, 2020, https://arxiv.org/abs/2003.07704.

[14]. Shouvanik Chakrabarti, Yiming Huang, Tongyang Li, Soheil Feizi, Xiaodi Wu. "Quantum Wasserstein GANs", arxiv, 2019, https://arxiv.org/abs/1911.00111.

[15]. Mehdi Mirza, Mehdi Mirza. "Conditional Generative Adversarial Nets", arxiv, 2014, https://arxiv.org/pdf/1411.1784.

[16]. WEI WANG, CHUANG WANG, TAO CUI, AND YUE LI. "Study of restrained network structure for Wasseerstein Generative Adverserial Natwork", IEEEAccess, 2020, Vol 10.1109, 10.1109/ACCESS.2020.2993839.

[17]. Ramin Hasani, Mathias Lechner, Alexander Amini, Daniela Rus, Radu Grosu. "Liquid Time- constant Networks", arxiv, 2020, https://arxiv.org/pdf/2006.04439.

[18]. HyunBum Kim, Juhyeong Jeon, Yeon Jae Han, YoungHoon Joo, Jonghwan Lee, Seungchul Lee, Sun Im. "Convolutional neural network classifies pathological voice change in laryngeal cancer with high accuracy", JCM, 2020, Vol 9(11), https://doi.org/10.3390/jcm9113415.

[19]. Balamurali B. T.; Hwan Ing Hee; O. H. Teoh; K. P. Lee; Saumitra Kapoor; Dorien Herremans; Jer-Ming Chen. "Asthmatic versus healthy child classification based on cough and vocalised /ɑ:/ sounds", JASA, 2020, Vol 148, issue 3, https://doi.org/10.1121/10.0001933.

[20]. Pauline Mouawad, Tammuz Dubnov, Shlomo Dubnov. "Robust Detection of COVID-19 in Cough Sounds", SN Computer Science, Vol 2, Article 34, 2021, https://doi.org/10.1007.