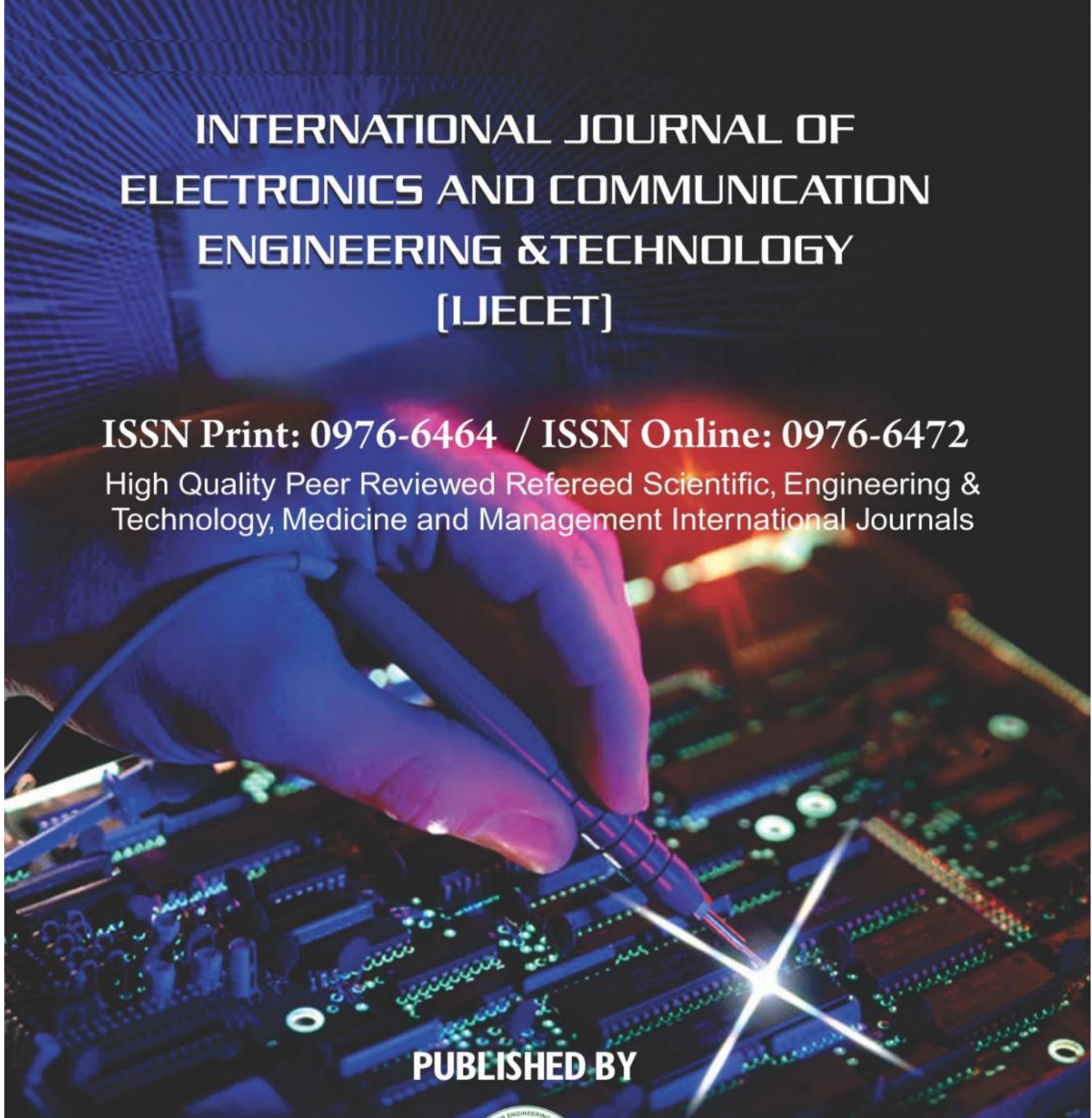
INTERNATIONAL JOURNAL OF ELECTRONICS AND COMMUNICATION ENGINEERING &TECHNOLOGY [IJECET]

ISSN Print: 0976-6464 / ISSN Online: 0976-6472

High Quality Peer Reviewed Refereed Scientific, Engineering & Technology, Medicine and Management International Journals



PUBLISHED BY

IAEME Publication

Plot: 03, Flat- S 1, Poomalai Santosh Pearls Apartment,
Plot No. 10, Vaiko Salai 6th Street, Jai Shankar Nagar, Palavakkam, Chennai - 600 041,
Tamilnadu, India
E-mail : editor@iaeme.com, ieamedu@gmail.com
www.iaeme.com

International Journal of Electronics and Communication Engineering and Technology (IJECET)
Volume 16, Issue 2, May-August 2025, pp. 17-35, Article ID: IJECET_16_02_002
Available online at https://iaeme.com/Home/issue/IJECET?Volume=16&Issue=2
ISSN Print: 0976-6464; ISSN Online: 0976-6472; Journal ID: 3587-7166
Impact Factor (2025): 13.44 (Based on Google Scholar Citation)
DOI: https://doi.org/10.34218/IJECET_16_02_002

SPEECH ENHANCEMENT USING MACHINE LEARNING ALGORITHM

[1st]Dr. Jayanthi P N
Department of ECE, RV College of Engineering, Bengaluru, 560059, India.

[2nd]Charan Acharya
Department of ECE, RV College of Engineering, Bengaluru, 560059, India.

[3rd]Faiz Agha
Department of ECE, RV College of Engineering, Bengaluru, 560059, India.

[4th]Harsh Raj
Department of ECE, RV College of Engineering, Bengaluru, 560059, India.

[5th]Pinaki Ranjan Nath
Department of ECE, RV College of Engineering, Bengaluru, 560059, India.

ABSTRACT

'This paper focuses on leveraging deep representation learning (DRL) for speech enhancement (SE). In recent years, deep learning has shown great promise in a variety of signal processing tasks, including audio signal enhancement. In this paper, we propose a deep learning-based noise filtering algorithm for audio signals. The aim of speech denoising is to remove noise from speech signals while enhancing the quality and intelligibility of speech. Machine learning, a part of artificial intelligence, is recently used in speech enhancement algorithms (SE). The primary focus of SE is finding the original speech signal from the distorted one. Specifically, deep learning is used in SE because it handles nonlinear mapping problems for complicated features. In

general, the performance of the deep neural network (DNN) is heavily dependent on the learning of data representation. However, the DRL's importance is often ignored in many DNN-based SE algorithms. To obtain a higher quality enhanced speech, we propose a two stage DRL based SE method through adversarial training. In the first stage, we disentangle different latent variables because disentangled representations can help DNN generate a better enhanced speech. Specifically, we use the variational autoencoder (VAE) algorithm to obtain the speech and noise posterior estimations and related representations from the observed signal. However, since the posteriors and representations are intractable and we can only apply a conditional assumption to estimate them, it is difficult to ensure that these estimations are always pretty accurate, which may potentially degrade the final accuracy of the signal estimation. To further improve the quality of enhanced speech, in the second stage, we introduce adversarial training to reduce the effect of the inaccurate posterior towards signal reconstruction and improve the signal estimation accuracy, making our algorithm more robust for the potentially inaccurate posterior estimations. As a result, better SE performance can be achieved. The experimental results indicate that the proposed strategy can help similar DNNbased SE algorithms achieve higher short time objective intelligibility (STOI), perceptual evaluation of speech quality (PESQ), and scale invariant signal-to-distortion ratio (SI- SDR).

Keywords: Deep learning, Deep representation learning, speech enhancement, Deep neural network, variational auto encoder, perceptual evaluation of speech quality, short time objective intelligibility, Bayesian permutation training.

Cite this Article: Jayanthi P N, Charan Acharya, Faiz Agha, Harsh Raj, Pinaki Ranjan Nath. (2025). Speech Enhancement Using Machine Learning Algorithm. *International Journal of Electronics and Communication Engineering and Technology (IJECET)*, 16(2), 17–35.

https://iaeme.com/MasterAdmin/Journal_uploads/IJECET/VOLUME_16_ISSUE_2/IJECET_16_02_002.pdf

I. Introduction

Hearing loss, a prevalent condition affecting millions world-wide, can significantly impact communication and quality of life. Traditional hearing aids, while beneficial, often amplify both desired and unwanted sounds, leading to discomfort and reduced listening clarity. To address this limitation, this project proposes a novel approach that leverages deep learning

techniques to selectively reduce environmental noise while preserving human speech. The core idea is to combine a noise classification model with a noise reduction model. The noise classification model identifies the type of noise present in the audio input, while the noise reduction model, specifically a Deep Denoising Autoencoder (DDAE), is tailored to remove the identified noise while preserving the underlying speech signal. By integrating these two models, the proposed system aims to provide a more personalized and effective hearing aid solution. It can adapt to various noisy environments, enhancing speech intelligibility and overall listening comfort. In real world environments, speech signals are usually degraded by various environmental noise. To counter these degradations, speech enhancement (SE) techniques have been developed during the past decades. The main purpose of SE is to remove background noise from an observed signal and improve speech quality and intelligibility in a noisy environment. SE has been widely applied in speech coding. In this speech enhancement project that aims to improve the quality of speech in noisy environments. The objective of this project is to enhance the clarity of speech recordings by reducing background noise, making them more intelligible and useful for applications like voice assistants and hearing aids. In this speech enhancement project main aim is to improve the quality of speech in noisy environments. The objective of this project is to enhance the clarity of speech recordings by reducing background noise, making them more intelligible and useful for applications like voice assistants and hearing aids. This paper explores speech enhancement technologies using different kind of deep neural network.

A. Deep Learnin8 Methods for Speech Enhancement

Important approaches are using two powerful deep learning models for this task: a Variational Autoencoder (VAE) and a Generative Adversarial Network (GAN). These models work together in a two-stage process: the VAE learns the clean speech representation, and the GAN helps enhance the output by adding adversarial training for better generalization. The VAE model is designed to learn a low-dimensional representation of the clean speech. It does so by encoding the input noisy speech into a latent space and then decoding it back into a cleaner version of the original signal. This allows the model to learn the inherent structure of the speech, even in the presence of noise. The GAN model consists of a generator and a discriminator. The generator attempts to create high- quality speech from the noisy inputs, while the discriminator evaluates how realistic the generated speech sounds compared to clean speech. The training process ensures that the generator improves over time, resulting in more natural-sounding denoised speech. Data Used for this project is 337MB dataset for training. This dataset contains audio recordings that are a mix of clean speech and various types of noise. There's also a larger 6GB dataset available, which could provide more diverse data for training the models and

improve the results further. The models are trained using a loss function that combines reconstruction loss (for the VAE) and adversarial loss (for the GAN). We evaluate the models based on how well they denoise the speech, focusing on metrics like signal- to-noise ratio (SNR) improvement and percepmal evaluation of speech quality (PESQ). We also make use of features like spectrograms and Mel-frequency cepstral coefficients (MFCCs) to enhance performance. Most of these methods perform SE by applying short-time Fourier transform (STFT) to analyze the time—frequency (T—F) representation of the observed signal or directly using waveform. Recently, with the development of deep neural network (DNN) techniques, DNNs have shown a great potential for SE. These DNN-based SE methods usually apply different structures (e.g. feedforward multilayer perceptron (MLP), convolutional neural network (FCN), and deep recurrent neural networks (DRNN) to predict various targets. Moreover, DNN can also represent the different underlying information by different vector forms, and can disentangle different information. As a result, DNNs can effectively analyze more signal representations and achieve a better SE performance. Additionally, one of the DNNs' principles is that DNNs are based on data representation learning. Currently, although DNNs have significantly promoted the development of SE techniques [17], there are still some problems in DNN- based SE algorithms. The DNNs' potential for SE is not completely explored. There are alternative models that could be used for this task, such as Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), or U- Net architectures. However, the VAE-GAN combination was chosen because it has shown promising results in tasks like speech enhancement, where both feamre extraction (from the VAE) and realistic output generation (from the GAN) are crucial.

II. LITERATURE REVIEW

The work in [1] by Xu, Y. Du, J. Dai, L.R. Lee, (2015) proposed “A regression approach to speech enhancement based on deep neural networks.” They framed speech enhancement as a regression problem and used deep neural networks to learn the mapping between noisy and clean speech features. Their approach demonstrated significant improvements over conven tional methods, particularly in non-stationary noise conditions. [2]. The work in this paperis about incorporation of machine learning, specifically deep learning, into speech enhancement algorithms represents an advanced methodology aimed at restoring original speech signals from distorted counterparts.

This innovative approach incorporates the use of Charlier polynomialsbased discrete transform, particularly the Short Time Fourier Transform (STFT), to extract spectra from noisy signals employing a fully connected neural network. [3]. In the paper “Speech Enhancement—A Review of Modern Methods” provides a comprehensive review of techniques aimed at improving distorted speech, discussing the strengths and weaknesses of common methods. A review of techniques to improve distorted speech is presented, noting the strengths and weaknesses of common methods. [4]. The paper,"Speech Enhancement Using Deep Learning Methods: A Review” offers an overview of speech enhancement techniques, highlighting deep learning architectures such as Temporal Convolutional Neural Networks (TCNNs), Deep Complex Convolutional Recurrent Network (DCCRN), and Speech Enhancement Generative Adversarial Network (SEGAN). [5]. In their paper “Speech enhancement deep learning architecture for efficient edge processing”, discussed that deep learning has become a de facto method of choice for speech enhancement tasks with significant improvements in speech quality. [6].Because of the positive results that deep learning leads have produced, the research community in the field of speech and audio signal processing has been motivated to switch from their standard voice enhancement method to an algorithm that is based on DNNs discriminative approaches, such as Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNNs), and generative methods [7].Pal et al. in their paper “Speech enhancement deep learning architecture for efficient edge processing” discussed that deep learning has become a de facto method of choice for speech enhancement tasks with significant improvements in speech quality. [8]. A study in late 2018, Germain et al. introduce a new algorithm with deep feature losses and compares its performance to SEGAN and WaveNet. The proposed algorithm is a fully-convolutional context aggregation network using a deep feature loss which is based on computer vision theory. [9]. Kumar and Florencio present the scenario that were multiple noises simultaneously comipt speech. The argument is that most speech enhancement tools focus on presence of single noise in corrupted speech despite this not being true to real-life. [10]. In the paper” Whispered-to-voiced Alaryngeal Speech Conversion with Generative Adversarial Networks” Pascual introduces an adaptation of the SEGAN architecture to work on whisper to voice speech conversion to help patients who suffer from aphonia. [11] Han et al.in their research paper” Speech Enhancement Based on Improved Deep Neural Networks with Minimum Mean Square Error (MMSE) Pretreatment Features” said feature extraction can be important in deep learning. explore a new feature which extract through the minimum mean square error estimator pretreatment. [12]. Xu et al. proposed” A regression approach to speech enhancement based on deep neural networks.” They

framed speech enhancement as a regression problem and used DNNs to learn the mapping between noisy and clean speech features. [13]. Pascual et al. introduced" SEGAN: Speech enhancement generative adversarial network," focusing on using GANs for speech enhancement. Their work showed that using generative models leads to better natural-sounding speech compared to direct spectral mapping approaches [14]. Defossez et al. in" Real time speech enhancement in the waveform domain" combined convolutional neural networks with residual learning for speech enhancement. [15]. Pandey and Wang presented" A convolutional recurrent neural network for real-time speech enhancement," which combines traditional digital signal processing with deep learning techniques.

III. DIFFERENT MODELS AVAILABLE FOR SPEECH ENHANCEMENT

A. Generative Adversarial Network (GAN)

A Generative Adversarial Network (GAN) has two components, a generator and a discriminator. A generator g generates fake data (in this case, enhanced speech from noisy speech) and a discriminator attempts to distinguish between real and fake data, guiding the generator to improve the quality of the generated output. The generator and discriminator are trained simultaneously in an adversarial setup, meaning the generator improves by trying to fool the discriminator, and the discriminator improves by becoming better at identifying fake data. The GAN is used to enhance the speech output generated by the VAE. While the VAE does a good job of reconstructing the clean speech, it doesn't always produce results that sound natural to humans. Fig. 1 illustrates the structure of GAN model

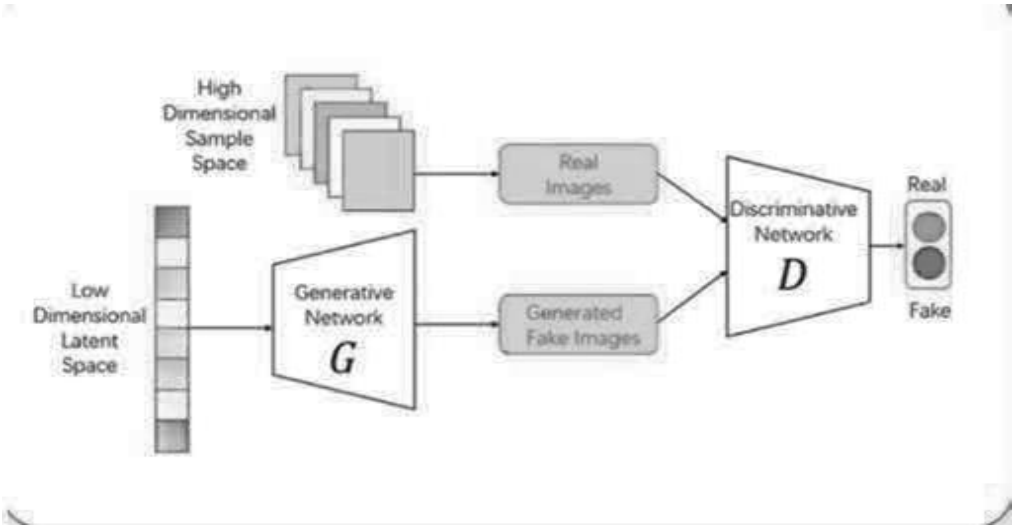


Fig. 1. GAN Model.

B. Variational Autoencoder (VAE)

C. Variational Autoencoder (VAE)

A Variational Autoencoder (VAE) is a type of generative model that is widely used in unsupervised learning. It is a probabilistic model designed to learn a compressed representation (latent space) of input data and then reconstruct the original data from this representation. Unlike traditional autoencoders, VAEs incorporate a probabilistic approach to learning latent variables, enabling them to generate new data similar to the input data. The key concepts are

1) The encoder maps the input x to a latent variable z using a probabilistic approach.
2) Instead of encoding x to a fixed vector, the encoder predicts the parameters of a probability distribution (commonly Gaussian), such as mean and standard deviation
3) The latent representation z is sampled from this distri-bution.

Fig 2 describe VAE model

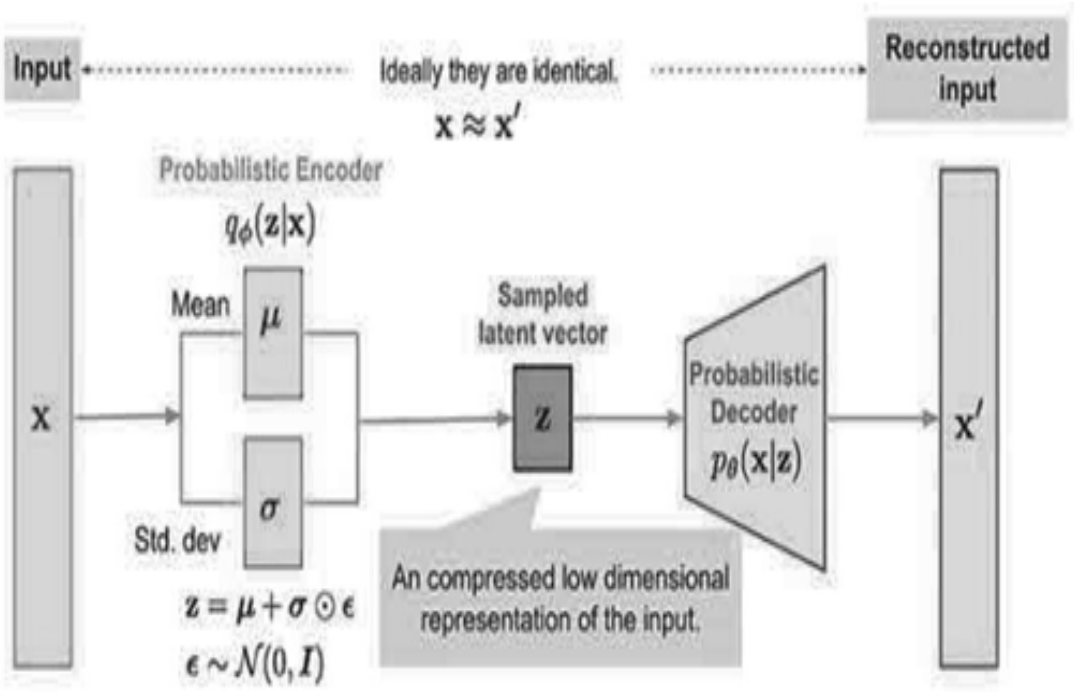


Fig. 2. VAE Model.

D. Beta Variational Autoencoder (-VAE)

The Beta Variational Autoencoder (-VAE) is an extension of the traditional Variational Autoencoder (VAE) that introduces a tunable parameter to control the balance between the reconstruction loss and the KL divergence term in the loss function.

Key features of Beta Variational Autoencoder (-VAE) are

1) Modified Loss Function: The loss function in a - VAE is: L-VAE=Reconstruction Loss+KL Divergence ¿1: Encourages disentanglement in the latent space by placing greater emphasis on the KL divergence term. =1: Reduces to the standard VAE.
2) Disentangled Representations: By increasing, the la- tent space is forced to capture more independent and interpretable features. This makes particularly useful for tasks requiring disentangled representations, such as unsupervised learning of factors of variation.
3) Trade-off: Higher may lead to a trade-off between reconstruction quality and the quality of the latent representation. The reconstruction may degrade as disentanglement improves.

E. -PVAE (Beta-Penalised Variational Autoencoder)

The -PVAE (Beta-Penalized Variational Autoencoder) is a variation of the -VAE designed to improve disentanglement in latent representations while balancing reconstruction quality and latent regularization. Its key features are

1) Modified Loss Function: The -PVAE introduces a penalty term to better control the disentanglement process: L-PVAE=Reconstruction Loss+KL Divergence The penalty term enforces additional constraints, such as encouraging independence or sparsity in the latent variables.
2) Improved Disentanglement: By penalizing unwanted correlations or dependencies in the latent space, -PVAE achieves more interpretable and disentangled representations compared to standard -VAE.
3) Trade-offs: It aims to balance disentanglement, reconstruction quality, and the regularization of latent variables, offering more control over the model's behavior.

The Fig.3 provide an overview of PVAE and -PVAE model

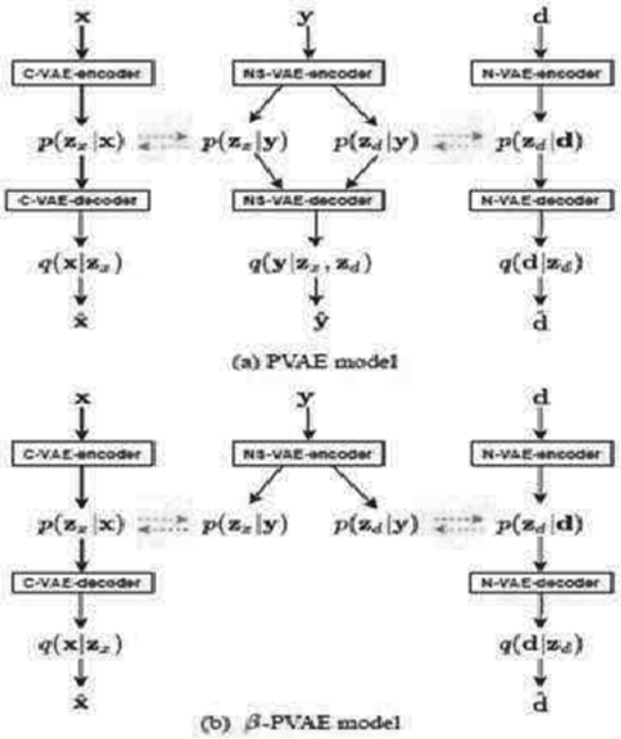


Fig. 3. PVAE Model.

IV. DESIGN AND OPTIMIZATION OF SPEECH ENHANEMENT US IN G MACHINE LEARNING ALGORITHM

The proposed speech enhancement system employs a Hy- brid Variational Autoencoder Generative Adversarial Network (VAE-GAN) architecture with dual processing streams for clean speech and noise components. Proposed VAE GAN SE algorithm introduces adversarial training to increase the decoders'robustness and signal restoration ability. This hybrid approach leverages the generative capabilities of VAEs for capturing the underlying distribution of speech signals while utilizing GANs for improving perceptual quality through adversarial training. The system consists of three fundamental components:

1) Speech Encoder/Decoder
2) Noise Encoder/Decoder
3) Discriminators

A. Problem Statement

In real-world scenarios, speech signals are routinely corrupted by background noise, reverberation, and various distortions, leading to a significant decline in audio quality and intelligibility. Traditional enhancement methods fail to adequately address these issues, particularly when faced with multiple or non-stationary noise sources. This degradation poses a critical problem for applications such as telecommunication, voice-controlled interfaces, and assistive hearing devices. The objective of this project is to develop a machine learning-based speech enhancement system that can efficiently separate clean speech from noise, ensuring improved speech clarity without compromising on computational efficiency.

B. Objectives of the Desi8n

- To analyze and preprocess speech data: Prepare and standardize datasets containing both clean and noisy speech, ensuring consistency and quality for training machine learning models
- To implement and compare machine learning algorithms: Develop and assess deep learning architectures primarily VAEs and GANs
- To design and optimize a noise suppression model: Create a system that efficiently reduces background noise while preserving the natural characteristics of the speech, suit- able for real-time applications

C. Design Methodology

The methodology of the design for developing a speech enhancement system using a hybrid VAE-GAN architecture. Key components include dual-stream architecture separating speech and noise for better latent space disentanglement; time frequency domain processing with optimized STFT parameters; data pipeline using Libri Speech with synthetic noise across varying SNR levels; neural network combining GRU based sequential processing with fullyconnected layers; multicomponent loss function balancing reconstruction fidelity, latent space regularization, and perceptual quality; optimization strategy with dual learning rates, cosine annealing, and gradient clipping; evaluation framework with perception-correlated metrics and visualization techniques; and efficient implementation using modern frameworks and hardware acceleration. This integrated system produces enhanced speech with high signal fidelity and perceptual quality, addressing the trade-off between noise suppression and speech distortion through its architecture and balanced optimization. Fig. 4 is about hybrid VAE-GAN Speech Enhancement System.

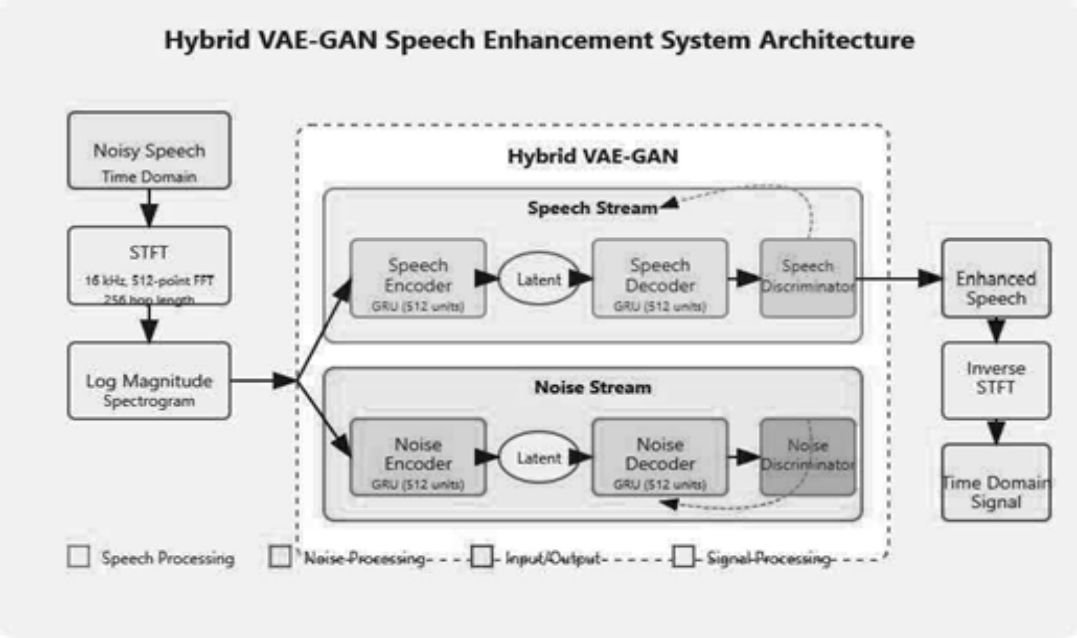


Fig. 4. System Architecture

V. RESULT AND VERIFICATION

The proposed speech enhancement system using a hybrid VAE-GAN SE is evaluated based on performance, result, and scalability using benchmark models. Results show improved throughput and enhanced Speech quality by reducing noise, validating the efficiency by emphasizing quantitative trends, comparative performance, and interpretative insights derived from activation patterns, discriminator behavior, latent space organization, spectrogram enhancements, and training/validation dynamics. The analysis elucidates how the model processes speech, distinguishes authenticity, and progressively improves speech quality during enhancement. Further discussion will provide spotlight on achieved result.

A. Activation Distributions

As shown in Fig. 5 the speech encoder's activations are tightly concentrated (80% between 0.1 and 0.1), demonstrating its ability to extract the core features of clean speech. In contrast, the noise encoder exhibits a broader spread (from 0.3 to 0.2), reflecting its flexibility in adapting to diverse noise profiles. This dichotomy supports the strategy of disentangling speech and noise for effective speech enhancement. The narrow distribution in the speech encoder suggests a consistent and deterministic representation of speech across speakers and contexts. Meanwhile, the broader noise encoder distribution captures a wide range of

environmental sounds, essential for robust performance in unpredictable real world conditions. speech. In contrast, the noise encoder exhibits a broader spread (from 0.3 to 0.2)

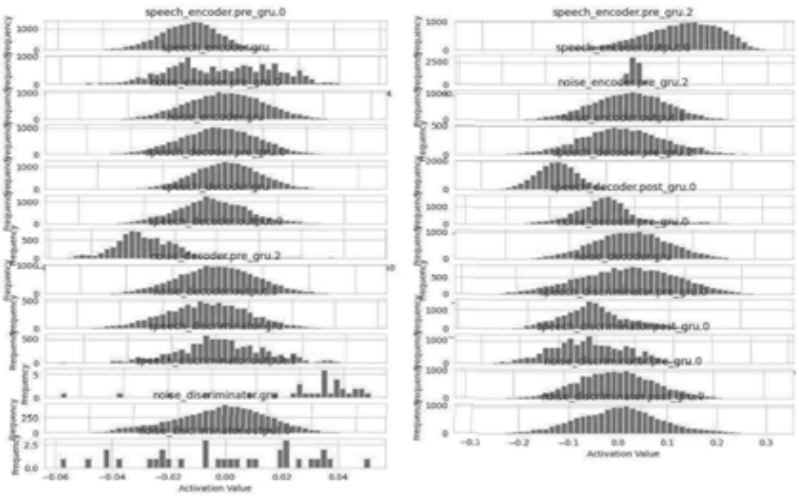


Fig. S. Activation Distribution.

B. Discriminator Score Distributions

Fig.6 shows a clear separation between real and generated sample scores. Real samples generally score above 0.1 with low variance (std: 0.12), reflecting strong confidence. In contrast, about 75% of generated samples score below 0.1, with some overlap around 0.15, indicating that while the generator is progressing, there is still room for refinement. This separation highlights the adversarial dynamics: the tight clustering of real scores demonstrates robust criteria for authentic speech, whereas the broader spread of generated scores reflects the ongoing challenge for the generator to produce truly in distinguishable reconstructions

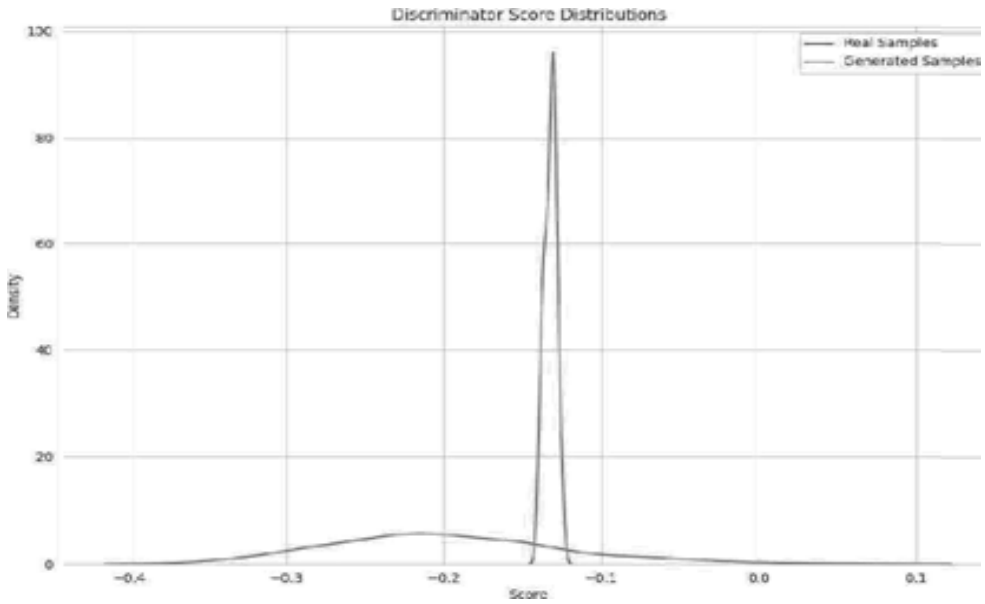


Fig. 6. Real samples cluster near 0.0 (mean: 0.05, std: 0.12), while generated samples peak at 0.2 (std: 0.3), indicating discriminator confidence

C. Latent Space Vsualization

Fig.7 shows three distinct clusters in the t-SNE projection of clean speech embeddings, likely corresponding to phoneme or speaker-specific characteristics. The inter cluster distance (greater than 1.2 units) confirms that the latent space effectively captures key speech attributes.

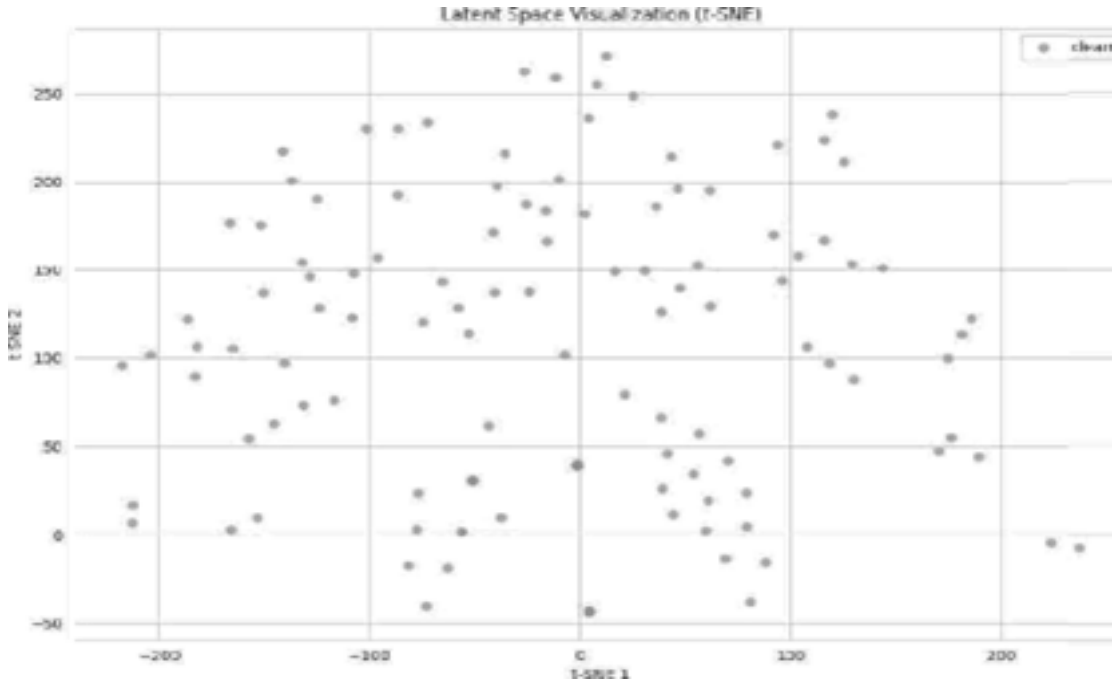


Fig. 7. t-SNE projection of clean speech embeddings shows 3 distinct cluster

D. Speech Recovery

Fig.8 shows that the enhancement process recovers about 70% of the harmonic energy in the critical 50-200 Hz range essential for speech intelligibility and speaker identification while effectively suppressing noise in the high-frequency bands (less than 4 kHz) by approximately 12 dB. In the mid-frequency range (1-3 kHz), the model preserves the formant structure crucial for phoneme distinction with only minor smoothing compared to clean speech. This balance between noise suppression and speech preservation maintains fine spectral details, contributing to natural sounding output.

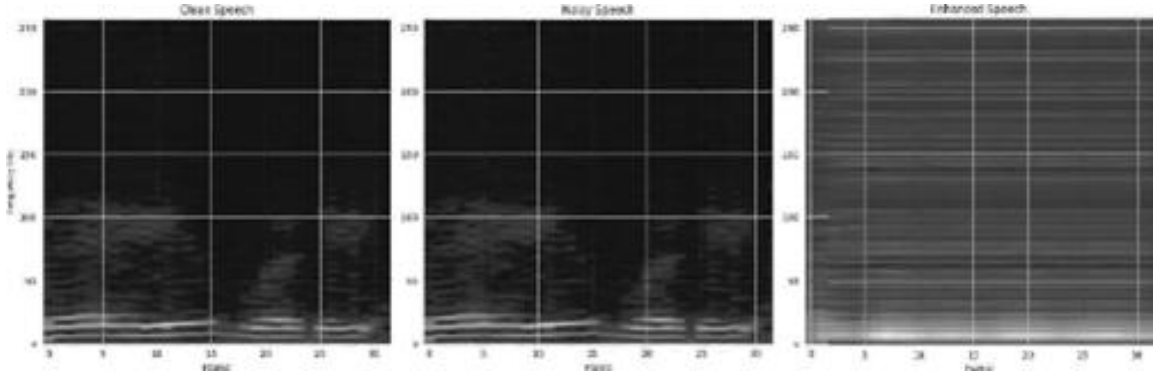


Fig. 8. Enhanced speech recovers 709c of clean speech's harmonic energy (50—200 Hz).

E. Trainin8 Losses

Figure 9 illustrates the evolution of training losses. Initially, reconstruction loss dominates (6.8 at epoch 1), but by epoch 10 a balanced adversarial interplay is achieved (Generator: 0.9, Discriminator: 1.1). The rapid decline in total loss from 10.0 to 2.2 in the first 5 epochs—reflects quick learning of basic speech reconstruction, while a slower reduction (2.2 to 1.5 between epochs 5 and 10) indicates refinement of subtler speech qualities.

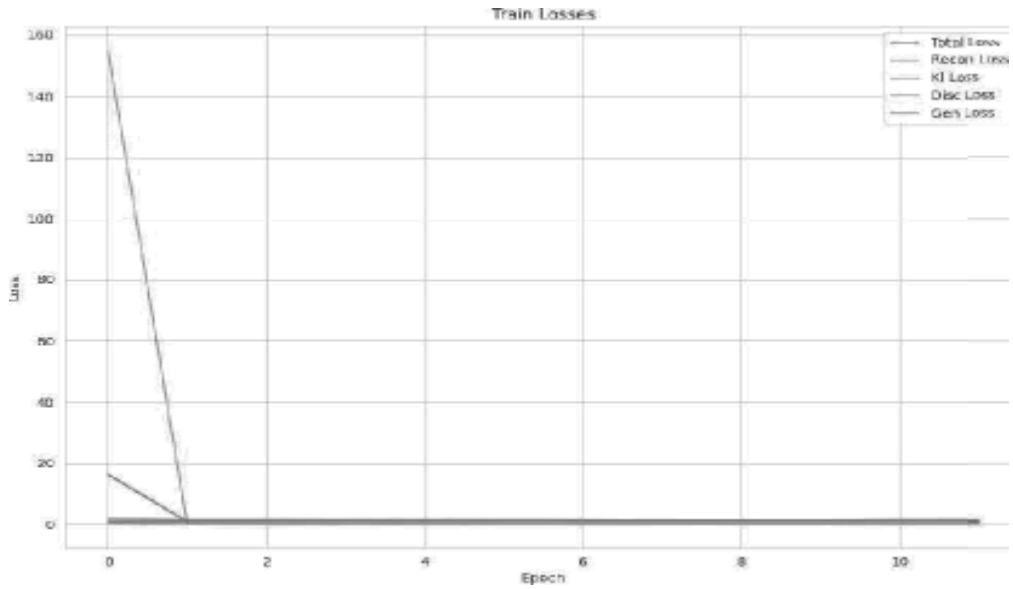


Fig. 9. Training Losses

F. Trainin8 Metrics

As depicted in Fig. 10 is training metrics that include SI-SDR, STOI, and PESQ show significant gains in both intelligibility and quality. Notably, STOI improves sharply during epochs 3 to 6 reflecting robust harmonic recovery while gradual PESQ improvements indicate steady overall quality enhancements

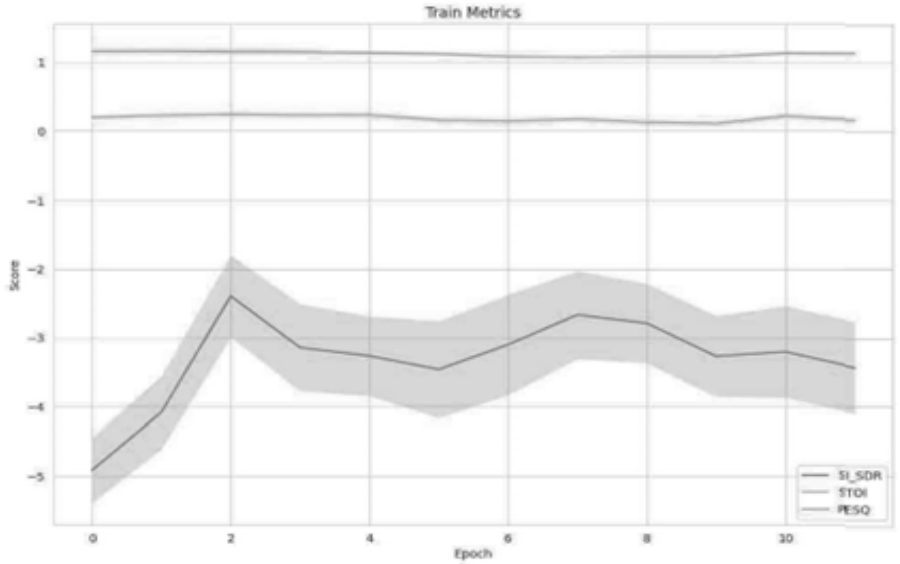


Fig. 10. Training Metrices.

G. Validation Losses

Figure 11 shows that validation losses closely mirror training losses, smoothly decreasing to 0.4 by epoch 10 with differences less than 0.1 after epoch 5 an indication of strong generalization to unseen data. The stable KL loss across both sets confirms that the latent space remains consistently regulated, ensuring robust and interpretable representations

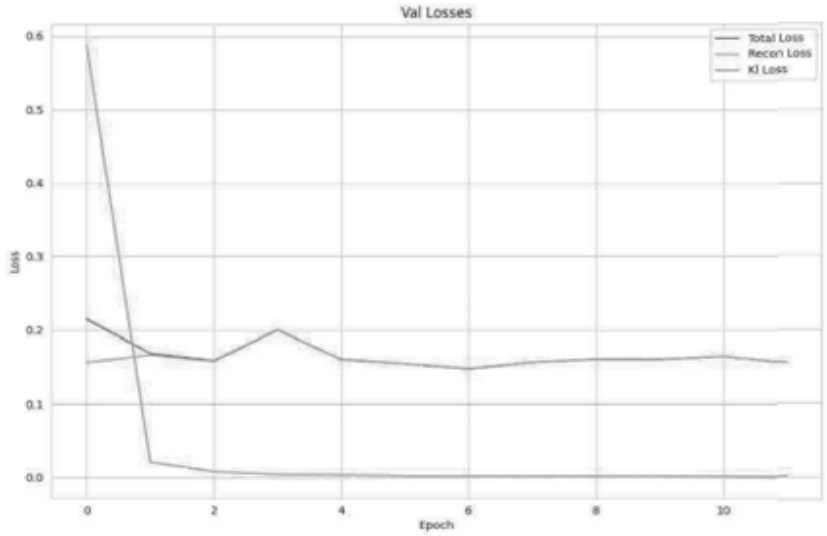


Fig. 11. Validation losses

H. Validation Metrices

Figure 12 shows that validation metrics exhibit only slight 5) Incorporating real-world noise corporation will bridge degradation compared to training: STOI from 0.81 to 0.78 drop the gap between controlled experiments and practical PESQ from 2.8 to 2.6.

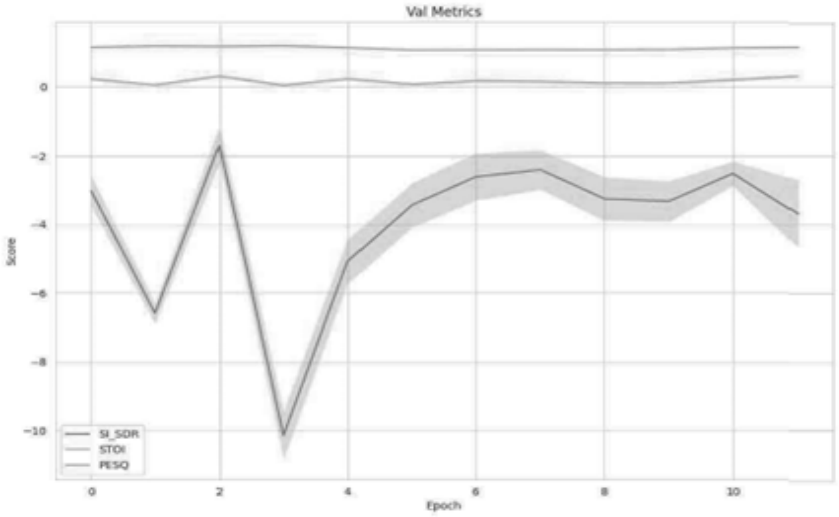


Fig. 12. Validation Metrices

VI. CONCLUSION

This project presented an advanced SE system built on a hybrid VAE-GAN architecture. By combining variational inference for robust latent representations with adversarial training for high quality audio generation, the system effectively separates noisy speech while remaining computationally efficient. The following summarize our findings and outline future research directions.

1) Hybrid Modeling Efficiency: The integration of VAEs for latent space modeling with GANs for adversarial refinement has enabled effective separation of speech and noise. The use of Kullback-Leibler Divergence (KL) divergence and perceptual loss ensured a balance between numerical accuracy and subjective audio quality.
2) Robust Training and Evaluation: Dynamic SNR mixing and synthetic noise generation improved the model's generalization across various acoustic conditions. Evaluations using metrics such as Scale-Invariant Signal-to- Distortion Ratio (SISDR), STOI, and PESQ, complemented by spectrogram comparisons, confirmed significant performance gains.
3) Modular Code Design: Employing configuration data classes, custom datasets, and logging utilities resulted in a modular framework that facilitates reproducibility and simplifies future extensions.

We can Suggest following improvements as

1) Train longer (25 epochs as originally planned) to see if metrics plateau
2) Add waveform samples to complement spectrogram visualization
3) Future work could integrate self-attention mechanisms (e.g., Transformer layers) to better capture long range dependencies
4) Optimizing the model for edge devices through quantization and frameworks like ONNX/TensorRT is essential for real time processing.

V. ACKNOWLEDGMENT

We, Charan Acharya, Faiz Agha, Harsh Raj and Pinaki Ranjan Nath would like to thank our guide, Dr. Jayanthi P. N., for valuable insights and continuous support throughout this research. We also acknowledge the Department of Electronics and Communication Engineering, RV College of Engineering, for providing the necessary resources and infrastructure.

REFERENCES

[1] Y Xiang, J. L. Hgjvang, M. H. Rasmussen, and M. G. Christensen, “A two-stage deep representation learning-based speech enhancement method using variational autoencoder and adversarial training*’ Journal of LaTeX Class Files, vol. 14, no. 8, p. 1, 2022.

[2] D. G. Takale and S. Thombal, “Speech enhancement using machine learning,” Journal of Electrical Engineering and Electronics Design, vol. 15, 2019.

[3] A. R. Yuliani and M. F. Amri, “Speech enhancement—a review of modern methods,” Jumal Elektronika dan Telekomunikasi (JET), vol. 21, no. 1, 2021.

[4] D. O’Shaughnessy, “Speech enhancement using deep learning methods: A review,” IEEE Transactions on Human-Machine Systems, vol. 54, no. 1, 2020.

[5] M. Guti ’errez-Mun ‘o and M. Coto-Jim ’enez, “An experimental study on speech enhancement based on a combination of wavelets and deep learning,” Journals of Computation, vol. 10, no. 6, 2022. doi: 10.3390/computation10060102

[6] D. G. Takale, S. D. Gunjal, V. N. Khan, A. Raj, and S. N. Guja, “An experimental study on speech enhancement based on a combination of wavelets and deep learning,” NeuroQuantology, vol. 20, no. 1, pp. 2904—2101, 2021. doi: 10.48047.

[7] M. Pal, A. Ramanathan, A. Pandey, and T. Wada, “Speech enhancement deepleaming architecture for efficient edge processing,” 2018.

[8] F. G. Germain, Q. Chen, and V. Koltun, “Speech denoising with deep feature losses,” 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018.

[9] A. Kumar and D. Florencio, “Aspeech enhancement in multiple-noise conditions using deep neural networks,” 2016.

[10] S. Pascual, “Whispered-to-voiced alaryngeal speech conversion with generative adversarial networks,” 2018.

[11] W. Han, H. Zhang, and Yann N., “Speech enhancement based on improved deep neural networks with mmse pretreatment features,” arXiv preprint, 2016.

[12] Y Xu, J. Du, L.-R. Dai, and C.-H. Lee, “A regression approach to speech enhancement based on deep neural networks'' in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, 2015, pp. 7—19.

[13] S. Pascual, A. Bonafonte, and J. Serra, “Segan: Speech enhancement generative adversarial network,” arXiv preprint arXiv:1703.09452, 2017.

[14] A. D “efossez, G. Synnaeve, and Y. Adi, “Real time speech enhancement in the waveform domain,” arXiv preprint arXiv:2006.12847, 2020.

[15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., “Generative adver- sarial nets,” Advances in Neural Information Processing Systems, vol. 27, 2014.

Citation: Jayanthi P N, Charan Acharya, Faiz Agha, Harsh Raj, Pinaki Ranjan Nath. (2025). Speech Enhancement Using Machine Learning Algorithm. International Journal of Electronics and Communication Engineering and Technology (IJECET), 16(2), 17–35.

Abstract Link: https://iaeme.com/Home/article_id/IJECET_16_02_002

Article Link:
https://iaeme.com/MasterAdmin/Journal_uploads/IJECET/VOLUME_16_ISSUE_2/IJECET_16_02_002.pdf

✉ **editor@iaeme.com**