CHAPTER 7
CONCLUSION

This research work attempts to contribute to speech enhancement, which is the initial step in many speech-related applications such as speech recognition, hearing aids, speaker recognition systems, Artificial Intelligence related speech applications and many more. Speech enhancement is improved when machine learning and deep learning algorithms are applied. Machine learning and deep learning algorithms are attempted in this research study.

The main focus of the research is single-channel speech enhancement, which is frequently subjected to additive background noise. The quality and intelligibility of speech need to be given importance as it aids in better understanding the listener. As the reduction of background noise has a direct impact on speech enhancement, the elimination of noise leads to improved speech perception.

More efforts are taken to implement machine learning and deep learning algorithms for speech enhancement. The algorithms used in this research are DFNN, Deep CNN, modified LSTM, and modified FCRN for noise reduction. The Deep CNN and modified LSTM give better enhancement results, and the best result is obtained from the modified FCRN technique. The quality and intelligibility measures demonstrate the performance evaluation in terms of speech enhancement.

The research becomes effective when it is applied to a community-based problem. Prosthetic devices produce a hoarse voice and include gurgling noise in the produced alaryngeal speech. The gurgling noise is produced due to the phlegm and mucus secreted in the throat and reaches the voice prosthesis. Speech enhancement with the four deep learning algorithms suggests that the quality and intelligibility of the speech produced can be improved. The algorithms are implemented to enhance the alaryngeal speech, and the enhanced alaryngeal speech is validated by analyzing the word error rate.

A unique resource that consists of speech sentences recorded from laryngectomy patients has been developed, which is the novelty of this research work. This could be used to develop algorithms to enhance the quality and intelligibility of speech for far-end speech enhancement systems.

The comparison of paralinguistic features indicates the difference in the features of alaryngeal speech. The MATLAB app helps to simulate real-world noise conditions at different noise levels.

7.1 SUMMARY

The most used conventional signal processing algorithms in speech enhancement are Spectral Subtraction, MMSE, and Iterative Wiener Filter. Considering traditional techniques like spectral subtraction, the main issue is that the improved speech has a distracting audible tonal feature. The practical implementations of the MMSE filter introduce far different suppression factors from the optimal. Compared to other traditional algorithms, the Iterative Wiener Filter performs moderately in improving speech quality and intelligibility. The hybrid approach seems better than the iterative Wiener filter in terms of the performance metrics.

As the speech signal is dynamic, these statistically-based unsupervised models are imperfect at predicting the variations. The statistical assumptions made by the unsupervised models need to be revised to improve the process of denoising the speech signals. The supervised models that are data-driven eliminate statistical assumptions concerning clean speech signals and noisy speech signals.

Currently, the enhancement techniques are based on the Artificial Intelligence taxonomy, where machine learning and deep learning techniques have become widely used to enhance intelligibility and improve speech's listening ability so that it is perceived effectively. The noise reduction is explicitly observed from the segSNR and SI-SDR values in the modified FCRN technique. The value of PESQ shows improved quality in modified LSTM and modified FCRN techniques for higher noise levels. Similarly, the STOI of modified FCRN is much improved compared to the other algorithms for all noise levels. The best algorithm in noise suppression is the modified FCRN technique, which is evident from the values of DNSMOS.

Speech enhancement helps improve speech quality for people who have undergone laryngectomy as their voice is very feeble and hoarse. Due to this reason, it is challenging for others to understand the sentences spoken by total laryngectomy patients with Blom Singer non-indwelling voice prosthesis. The WER is 100% when the alaryngeal speech is subjected to noise and is improved while applying speech enhancement algorithms. The

alaryngeal speech signal enhanced by the modified FCRN technique results in a lower word error rate and increases the chances of recognition of speech sentences. The research objective has been successfully achieved, demonstrating that the modified FCRN technique delivers superior results compared to other deep learning algorithms discussed in the research work.

7.2 SUGGESTIONS FOR FUTURE RESEARCH

The paralinguistic speech features such as pitch, tone, and frequency vary for voice produced by prosthesis when compared with the normophonic speech. Therefore, the gender of the voice cannot be discriminated against. The prosthetic device's construction could be re-engineered to produce a near-natural voice for both male and female speakers. The electronic prosthesis can be designed based on gender, and the generative AI based speech enhancement algorithm can be implemented so that the gurgling noise can be removed and the person's alaryngeal speech can be heard effectively.

Other prostheses like Groningen and Provox can be studied similarly, and a comparative evaluation could be made to assess the performance. Also, the pneumatic bionic voice technology Ahmadi et al. (2019) has a tremendous tangible opportunity to significantly enhance the quality of life for individuals with larynx amputation by restoring their communication ability.

.