INTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD
ISSN(O): 2455-0620 [ Impact Factor: 9.47 ]
Monthly, Peer-Reviewed, Refereed, Indexed Journal with IC Value : 86.87
Volume - 12, Issue - 8, August - 2026

DOIs:10.2015/IJIRMF/202608019 --:-- Research Paper / Article / Review

# CNN and DenseNet201-Based Speech Enhancement for Hearing-Assistive Applications

**[1] Dr. S. Swarnalatha, [2] D. Pooja**
[1] Professor, Department of Electronics and Communication Engineering, S V University College of Engineering, Tirupati, Andhra Pradesh, India
[2] Student, M.Tech, Department of Electronics and Communication Engineering, S V University College of Engineering, Tirupati, Andhra Pradesh, India
Email – [1] swarnasvu09@gmail.com, [2] Dandepooja438@gmail.com

***Abstract:*** *This study presents a comparative evaluation of a conventional convolutional neural network (CNN) and a DenseNet201-based deep convolutional model for single-channel speech enhancement in hearing-assistive applications. The approach uses short-time Fourier transform (STFT) analysis to derive a log-power-spectrum (LPS) representation from noisy speech, with corresponding clean-speech LPS features used as supervised targets. The predicted enhanced magnitude spectrum is reconstructed using the phase of the noisy signal, followed by inverse STFT and overlap-add processing. Performance is evaluated using segmental SNR, SNR improvement, SI-SDR, STOI, PESQ and log-spectral distance (LSD). In the reported experiments, DenseNet201 achieved higher SNR, SI-SDR, SNR improvement, STOI and PESQ than the conventional CNN across the two evaluated speech signals. However, DenseNet201 produced higher LSD, so improvement was not uniform across all spectral measures. The findings support dense feature reuse as a useful architectural strategy for LPS-based enhancement, while highlighting the need for larger held-out test sets, complete reproducibility details and computational profiling before real-time or clinical claims are made.*

***Key Words:*** *Speech Enhancement, Hearing Aid, Convolutional Neural Network, DenseNet201, Log Power Spectrum, Noise Reduction, Speech Quality, Speech Intelligibility, PESQ, STOI, SI-SDR.*

## 1. INTRODUCTION

Speech communication in real-world environments is frequently affected by background noise from traffic, competing speakers, machinery and public spaces. As the signal-to-noise ratio decreases, both speech quality and intelligibility can deteriorate. The problem is particularly important in hearing-assistive applications, where users may already have reduced access to speech cues. CNN-based enhancement has previously been investigated specifically for hearing-impaired listeners using smartphone platforms, demonstrating the relevance of deep-learning speech enhancement to assistive applications. [1]

Traditional enhancement methods such as spectral subtraction and Wiener filtering are computationally attractive but can introduce speech distortion and may be less robust to non-stationary noise. Deep learning provides a data-driven alternative in which the relationship between noisy and clean speech representations is learned from paired examples. Supervised enhancement has also been investigated for smartphone-based binaural hearing aids. [2]

Convolutional neural networks are attractive for spectral enhancement because convolution can learn local time-frequency patterns. CNN-based speech restoration and enhancement have been explored in different architectures, including recurrent-plus-CNN systems and temporal-envelope reconstruction. [3]

In this study, the existing project implementation is formulated as a comparative evaluation between a conventional shallow CNN and a DenseNet201-based CNN using log-power-spectrum features. DenseNet was originally introduced as a densely connected convolutional architecture that encourages feature reuse and improved information flow. [5]

The objective is to determine whether dense connectivity and feature reuse provide measurable benefits under the reported experimental conditions.

## 2. LITERATURE REVIEW

Bhat et al. demonstrated a real-time CNN-based speech-enhancement system intended for hearing-impaired listeners using a smartphone, establishing the relevance of CNN-based enhancement to hearing-assistive applications. [1]
Sun et al. investigated supervised speech enhancement for smartphone-based binaural hearing aids and emphasized speech quality and intelligibility under noisy conditions. [2]
Strake et al. proposed a two-stage deep-learning approach in which LSTM-based noise suppression was followed by CNN-based speech restoration. Soleymanpour et al. investigated CNN-based reconstruction of the speech temporal envelope in noisy environments. [3,4]
Densely connected convolutional networks were introduced by Huang et al., where each layer receives feature maps from preceding layers. This feature-reuse mechanism motivates the DenseNet201 comparison in the present study. [5]
The present project differs from highly complex multi-stage enhancement systems by using a supervised LPS mapping pipeline and comparing a conventional CNN with DenseNet201 under the same general preprocessing and reconstruction framework.

## 3. OBJECTIVES

- To develop a supervised LPS-based speech-enhancement pipeline for noisy single-channel speech.
- To compare a conventional CNN with a DenseNet201-based model under the same general enhancement framework.
- To evaluate enhancement using SNR, SNR improvement, SI-SDR, STOI, PESQ and LSD.
- To assess the suitability of the approach for hearing-assistive applications while explicitly identifying experimental limitations.

## 4. RESEARCH METHOD :

### • 4.1 Signal preprocessing

Clean and noisy speech signals are converted to a common single-channel representation. Each signal is segmented into short-time frames using a Hamming window and transformed using the STFT.

### • 4.2 Log-power-spectrum representation

The magnitude spectrum is squared to obtain the power spectrum. A logarithmic transformation is then applied to produce the LPS representation. Noisy and clean representations are aligned so that corresponding frames can be used as supervised input-target pairs.

### • 4.3 Conventional CNN

The baseline network is a shallow convolutional regression model designed to learn local spectral relationships. Convolutional layers with nonlinear activation functions extract spectral features and map noisy LPS features toward the corresponding clean LPS target.

### • 4.4 DenseNet201

DenseNet201 is used as the deeper convolutional model. Dense connectivity enables later layers to receive information from preceding layers, encouraging feature reuse and improving gradient propagation. The model is used to learn the mapping from noisy LPS features to clean LPS features. [5]

### • 4.5 Training

The models are trained in a supervised manner with noisy LPS features as input and clean LPS features as targets. The supplied project report specifies Adam optimization and a mean-squared-error objective.

### • 4.6 Speech reconstruction

The predicted LPS is converted back to a magnitude spectrum using the inverse logarithmic transformation. The phase of the original noisy speech is retained. The reconstructed complex spectrum is transformed back to the time domain using inverse STFT and overlap-add processing, followed by amplitude normalization.

### • 4.7 Evaluation metrics

The study reports segmental SNR, SNR improvement, SI-SDR, STOI, PESQ and LSD. Higher values are desirable for SNR, SNR improvement, SI-SDR, STOI and PESQ, whereas lower LSD indicates lower spectral distance. STOI is an intelligibility-oriented measure, PESQ is a perceptual speech-quality measure, and SI-SDR is an improved source-separation quality metric. [6,7,8]

## 5. FINDINGS / RESULTS

The numerical values below are reproduced from the supplied project report. Mean values are calculated from the two reported test signals. Because the original code and experiment logs are unavailable, these values should be treated as reported project results rather than independently reproduced measurements.

| Metric | CNN S1 | DenseNet201 S1 | CNN S2 | DenseNet201 S2 | Mean CNN | Mean DenseNet201 |
|---|---|---|---|---|---|---|
| SNR (dB) | 7.22 | 14.36 | 8.11 | 12.11 | 7.67 | 13.24 |

| | | | | | | |
|---|---|---|---|---|---|---|
| SI-SDR (dB) | 49.86 | 68.29 | 53.82 | 64.33 | 51.84 | 66.31 |
| SNR Improvement (dB) | 17.34 | 26.27 | 19.13 | 24.27 | 18.24 | 25.27 |
| LSD | 5.05 | 7.26 | 5.556 | 8.21 | 5.303 | 7.735 |
| STOI | 0.65 | 0.73 | 0.66 | 0.74 | 0.655 | 0.735 |
| PESQ | 3.50 | 4.20 | 3.15 | 3.50 | 3.325 | 3.85 |

*Table 1. Comparative performance reported for the two test signals.*

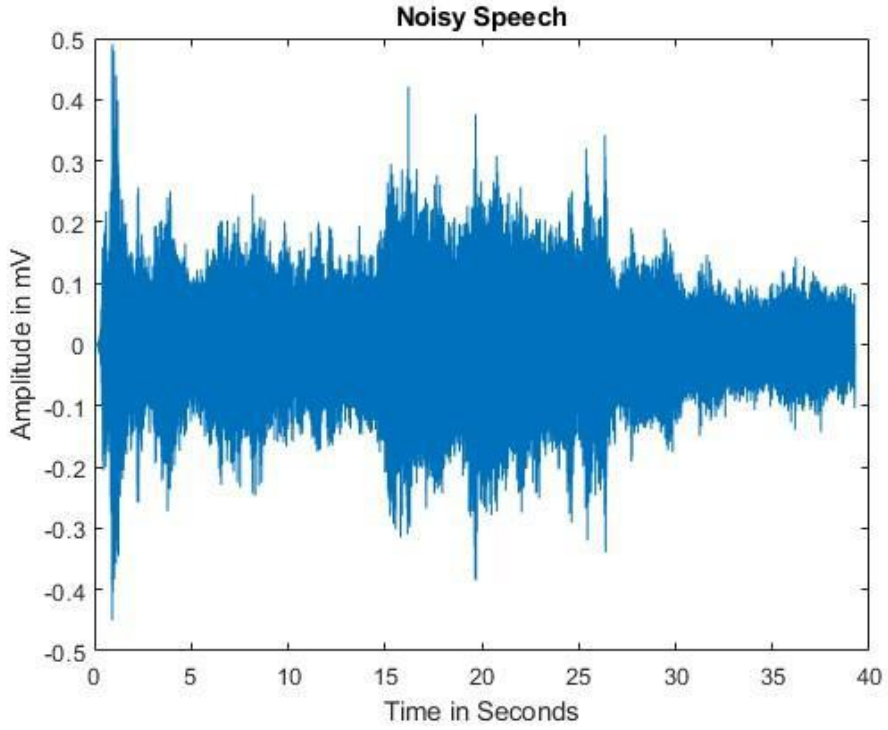


*Figure 1. Noisy speech waveform reported in the original project.*

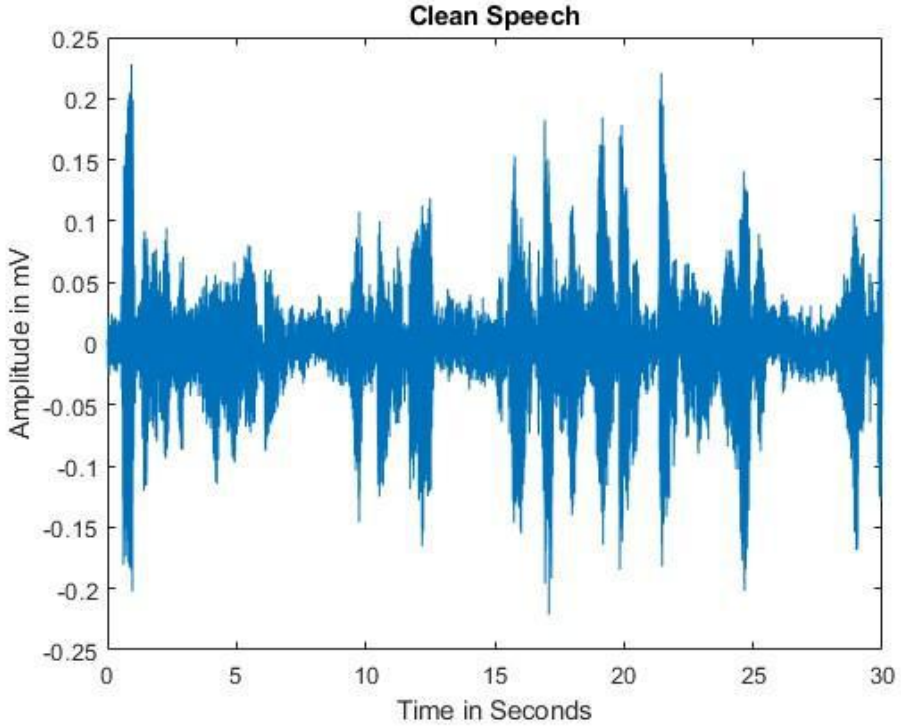


*Figure 2. Clean speech waveform reported in the original project.*

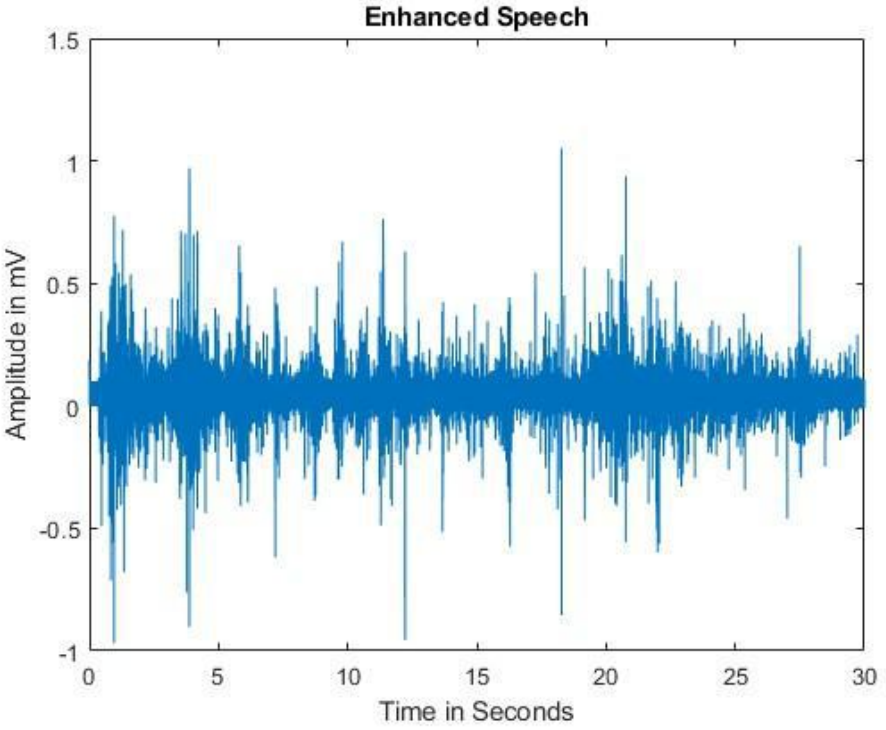


*Figure 3. Enhanced speech waveform produced by the reported method.*

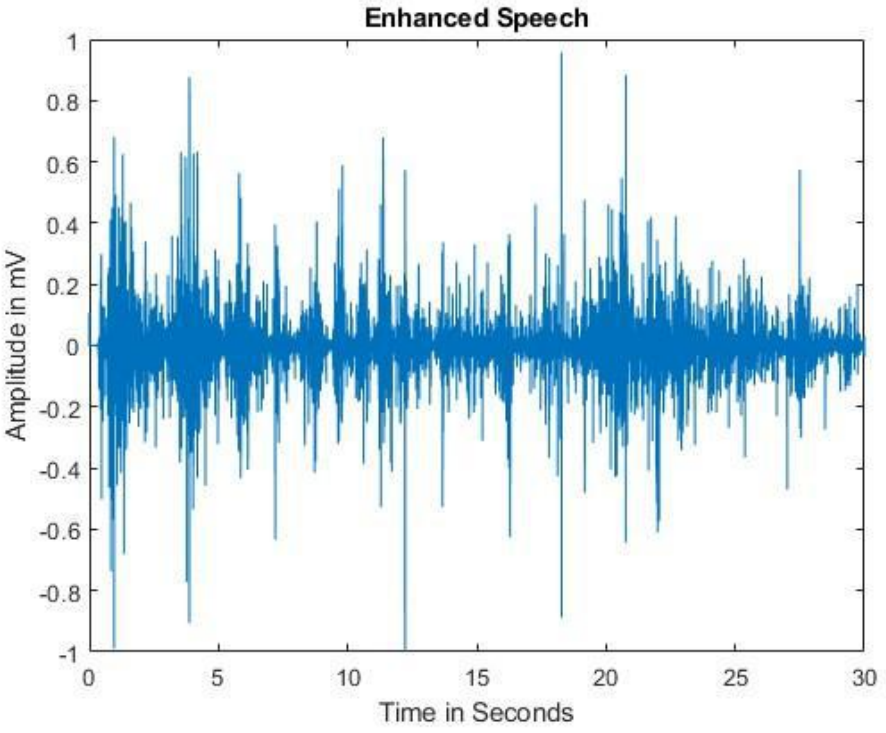


*Figure 4. Enhanced speech waveform for the second reported test signal.*

## 6. ANALYSIS

Across the two reported signals, DenseNet201 produced higher mean segmental SNR (13.24 dB versus 7.67 dB), SI-SDR (66.31 dB versus 51.84 dB), SNR improvement (25.27 dB versus 18.24 dB), STOI (0.735 versus 0.655) and PESQ (3.85 versus 3.33) than the conventional CNN.

The LSD result requires particular care. Mean LSD was 7.74 for DenseNet201 compared with 5.30 for the conventional CNN. Since lower LSD indicates lower spectral distance, DenseNet201 did not outperform the CNN on this measure. This demonstrates that enhancement quality is multidimensional and should not be inferred from a single metric.

The higher STOI and PESQ values suggest that DenseNet201 learned spectral representations that were more favorable to intelligibility and perceptual quality under the reported conditions. Dense connectivity may facilitate reuse of low-level spectral information while providing deeper feature extraction.
However, the evaluation contains only two reported test signals. The results should therefore be considered an initial comparative evaluation rather than a statistically generalizable benchmark. A larger held-out test set with multiple speakers, noise types and SNR conditions would provide stronger evidence.

## 7. CONCLUSION

This study presented a comparative evaluation of a conventional CNN and DenseNet201 for LPS-based single-channel speech enhancement intended for hearing-assistive applications. Under the reported two-signal evaluation, DenseNet201 achieved higher SNR, SI-SDR, SNR improvement, STOI and PESQ, while the conventional CNN achieved lower LSD.
The findings support the use of dense convolutional feature reuse as a promising strategy for spectral speech enhancement, but the limited evaluation prevents broad generalization. The present results should be interpreted as a project-level comparative evaluation rather than evidence of clinical efficacy or guaranteed real-time performance.

## 8. LIMITATIONS

- The supplied report does not provide the exact dataset size, train-validation-test split, number of epochs, batch size, learning-rate schedule, hardware configuration, model parameter counts or measured inference time.
- Only two test signals are reported in the quantitative comparison. No subjective listening evaluation involving hearing-impaired listeners is reported.
- The reconstruction stage retains the noisy phase, so the network primarily improves the magnitude/LPS representation. Phase-related distortion may therefore limit waveform quality.
- No latency, memory, power or real-time-factor measurements are available; consequently, real-time, lightweight and clinical-use claims are not established by the present evidence.

## 9. RECOMMENDATIONS

- Evaluate the complete held-out test set using multiple speakers, noise types and SNR levels.
- Report mean ± standard deviation and confidence intervals where the sample size permits.
- Measure parameter count, inference time, memory use and real-time factor for deployment-oriented claims.
- Investigate complex-spectrum or phase-aware prediction to reduce phase-related reconstruction limitations.
- Evaluate on representative smartphone or hearing-assistive hardware if deployment is a project objective.
- Conduct appropriate subjective intelligibility and quality testing with ethics approval where human participants are involved.

## 10. ACKNOWLEDGMENT

The authors acknowledge the guidance and academic support provided by the Department of Electronics and Communication Engineering, S V University College of Engineering, Tirupati, Andhra Pradesh, India.

## 11. CONFLICT OF INTEREST

The authors declare no conflict of interest in relation to this work.

## REFERENCES:

1. Bhat, G. S., Shankar, N., Reddy, C. K. A., and Panahi, I. M. S., “A real-time convolutional neural network based speech enhancement for hearing impaired listeners using smartphone,” IEEE Access, vol. 7, pp. 78421–78433, 2019. doi:10.1109/ACCESS.2019.2922370.
2. Sun, Z., Li, Y., Jiang, H., Chen, F., Xie, X., and Wang, Z., “A supervised speech enhancement method for smartphone-based binaural hearing aids,” IEEE Transactions on Biomedical Circuits and Systems, vol. 14, no. 5, pp. 951–960, 2020. doi:10.1109/TBCAS.2020.2988121.

3. Strake, M., Defraene, B., Fluyt, K., Tirry, W., and Fingscheidt, T., "Speech enhancement by LSTM-based noise suppression followed by CNN-based speech restoration," EURASIP Journal on Advances in Signal Processing, vol. 2020, art. 49, 2020. doi:10.1186/s13634-020-00707-1.
4. Soleymanpour, R., Soleymanpour, M., Brammer, A. J., Johnson, M. T., and Kim, I., "Speech enhancement algorithm based on a convolutional neural network reconstruction of the temporal envelope of speech in noisy environments," IEEE Access, vol. 11, pp. 5328–5336, 2023. doi:10.1109/ACCESS.2023.3236242.
5. Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q., "Densely connected convolutional networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4700–4708, 2017.
6. Taal, C. H., Hendriks, R. C., Heusdens, R., and Jensen, J., "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 7, pp. 2125–2136, 2011. doi:10.1109/TASL.2011.2114881.
7. Rix, A. W., Beerends, J. G., Hollier, M. P., and Hekstra, A. P., "Perceptual evaluation of speech quality (PESQ): a new method for speech quality assessment of telephone networks and codecs," Proceedings of ICASSP, pp. 749–752, 2001. doi:10.1109/ICASSP.2001.941023.
8. Le Roux, J., Wisdom, S., Erdogan, H., and Hershey, J. R., "SDR – half-baked or well done?" Proceedings of ICASSP, pp. 626–630, 2019.