Proceedings of the 5th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/100/2025.18285

MFCC-based Classification of Carotid Artery Doppler Audio Signals Using LSTM Network

Vidya Gopal T V[1], Inayathullah Ghori[2], Avinash Eranki[1], Renu John[1,a,*]

[1]Department of Biomedical Engineering, Indian Institute of Technology, Hyderabad, India
[2]Kamineni Hospital, King Koti, Hyderabad, India
a. renujon@bme.iith.ac.in
**corresponding author*

Abstract: The carotid artery assessment is essential for detecting stenosis and vascular abnormalities. Traditional Doppler ultrasound, while effective, requires specialized equipment and trained operators, limiting its accessibility in primary care. This study investigates Doppler audio signal analysis as a non-invasive, cost-effective alternative for assessing carotid artery hemodynamics. Using advanced signal processing techniques like mel-frequency cepstral coefficients (MFCCs) and deep learning models such as Long Short-Term Memory (LSTM) networks, we analyze Doppler audio signals from the common carotid artery (CCA) in 216 individuals. Our findings reveal significant age-related variations in blood flow dynamics and distinct signal patterns, highlighting the potential of Doppler audio analysis for early vascular screening. The changes in MFCCs indicate their usefulness in identifying hemodynamic alterations associated with aging and disease, supporting their role in non-invasive carotid artery health assessment. We also evaluate the deep learning framework, utilizing RNNs to capture long-term dependencies in the signals and providing a comprehensive comparison of network configurations and performance relative to state-of-the-art algorithms.

Keywords: Common Carotid Artery, Doppler Ultrasound Audio, Recurrent Neural Network, MFCC Features, Age related variations.

1. Introduction

The carotid arteries are vital for brain blood supply and are prone to vascular diseases like atherosclerosis and stenosis. Early detection and intervention are essential to prevent severe complications such as stroke. Traditional assessment methods, like Doppler ultrasound imaging, often require specialized equipment and skilled operators. Recent advancements in signal processing and machine learning have facilitated the analysis of Doppler audio signals, which capture the acoustic features of blood flow in the carotid arteries. By extracting key features from these signals, it is possible to evaluate vascular health non-invasively and cost-effectively. Advancements in signal processing and machine learning have facilitated the analysis of Doppler audio signals, allowing for non-invasive and cost-effective evaluation of vascular health. Advanced signal processing techniques, such as root mean square (RMS) and zero-crossing rate (ZCR), provide insights into basic signal characteristics, while frequency-domain techniques like Fast Fourier Transform (FFT) and Short-Time Fourier Transform (STFT) reveal spectral components associated with vascular abnormalities.

Time-frequency methods like wavelet transforms capture dynamic changes, offering deeper insight into transient events. Mel-Frequency Cepstral Coefficients (MFCCs) are highly effective in Doppler ultrasound analysis, mapping frequencies onto a Mel scale to extract spectral features linked to stenosis and atherosclerosis [1,2]. Integrating MFCCs with machine learning algorithms enhances diagnostic accuracy, providing noise-resistant, compact spectral representations [3,4].

2. Method

2.1. Data Collection

The study involved 216 participants aged 18-59 years without known vascular disorders, who provided written informed consent and provided a protocol approved by the institutional ethics committee. Doppler ultrasound data was obtained from the left and right common carotid arteries (CCA), carotid bifurcations (CB), and internal carotid arteries (ICA) using a wireless ultrasound system (FREESONO X-L4) with a 38 mm linear array transducer operating at 4-10 MHz. Participants lay supine with heads rotated 10-20 degrees for optimal placement, and each recording session lasted about 10 seconds [5,6]. Exclusion criteria included a history of vascular disease, brain trauma, stroke, or conditions compromising data integrity. Collected acoustic signals were saved in.wav format for analysis.

2.2. MFCC based signal processing

Preliminary preprocessing of raw Doppler audio data ensured its suitability for our application, focusing on the bandwidth of interest (0-2000 Hz). The MFCC reflect the nonlinear relationship between the human ear and sound frequency, expressed by the following relationship of the signal [7,8].

$$\text{Mel (f)} = 2595\log(1+\frac{f}{700}) \tag{1}$$

Where Mel(.) is the Mel scale frequency and F is the actual frequency of the signal.

Figure 1 shows the MFCC feature extraction process. The MFCC based signal processing reflected the nonlinear relationship between the human ear and sound frequency, expressed by the relationship of the signal. The first step involved framing the signal into equal sections (frames) with a typical duration of 25 ms and a hop length of 10 ms. A Hamming window was applied to each frame to preserve signal.

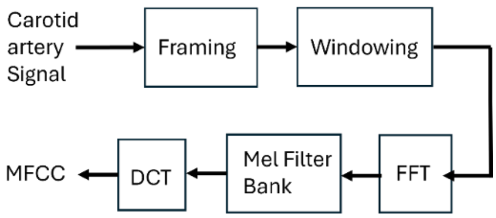


Figure 1: MFCC Feature Extraction

Next, a mel filter bank is applied to convert the frequency domain representation into a more perceptually meaningful form, capturing the non-linear characteristics of human hearing. Finally, the discrete cosine transform (DCT) is utilized to compute the most significant coefficients, yielding the MFCCs, which serve as the key features for further analysis.

2.3. The RNN classification model

This study proposes an RNN-based approach using Long Short-Term Memory (LSTM) layers to perform feature learning on MFCC representations derived from carotid artery audio signals. MFCC features offer a time-frequency representation that captures the unique characteristics of audio signals, which are processed by LSTM layers to exploit temporal relationships [9].

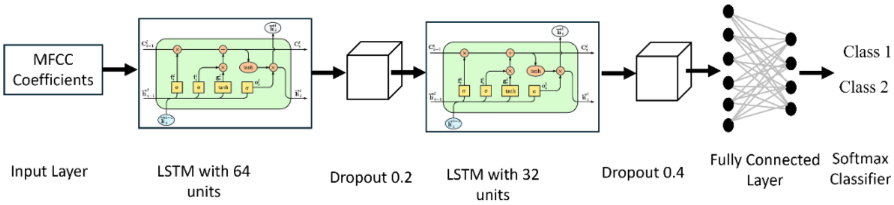


Figure 2: RNN Block Diagram

The proposed LSTM model consists of two layers: one with 64 units with L2 regularization to mitigate overfitting and a dropout layer to enhance generalization, and another with 32 units with L2 regularization and a 50% dropout layer. A fully connected layer with 32 units and ReLU activation is added to increase robustness. The output layer is a softmax layer that classifies input signals as normal or abnormal based on class probabilities. The RNN model leverages LSTMs' ability to handle sequential data, crucial for analyzing MFCC features from carotid artery audio signals.

3. Result

3.1. Distribution of center frequency on application of mel filter bank

We applied a mel filter bank to the Doppler audio signals to capture relevant spectral information. The 15 filters are spaced according to the non-linear mel scale, reflecting human auditory perception, with each filter emphasizing a specific frequency range and overlapping peaks to ensure smooth transitions between adjacent frequencies. This process is crucial for extracting MFCCs used in the analysis [10].

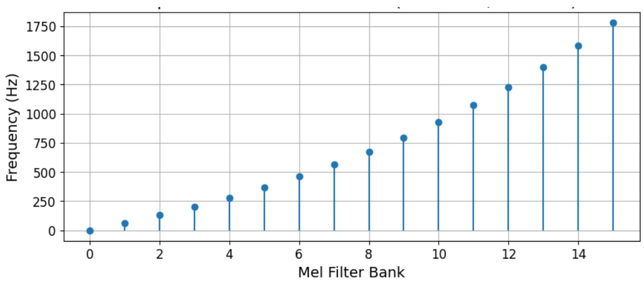


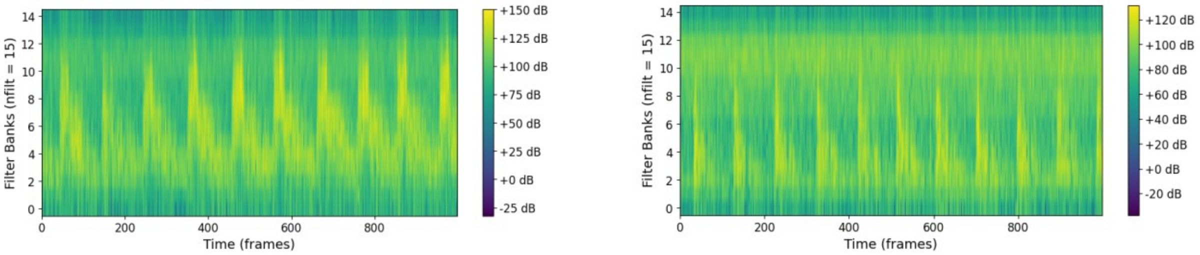
Figure 3: Distribution of center frequencies

Figure 3 illustrates the distribution of center frequencies for the first 13 filters, which are densely packed at lower frequencies (below 1000 Hz) and spaced wider at higher frequencies, aligning with the human ear's sensitivity.

DOI: 10.54254/2755-2721/100/2025.18285

3.2. Comparison of filter bank output and cepstral coefficients for younger and older samples

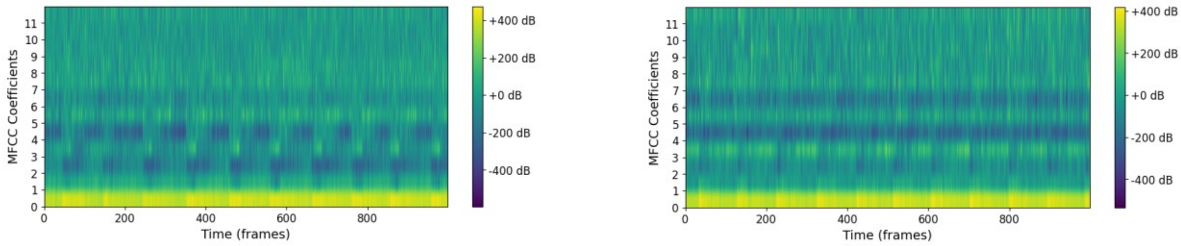
Filter The analysis reveals that the primary energy of the Doppler signals is concentrated in the lower filter banks (1–15), indicating that the most relevant spectral information is found in the lower-frequency range.



a. 24 years old b. 48 years old

Figure 4: Sample Spectrogram of Filter Bank Outputs

Figure 4a shows the signal from samples of younger age group with distinct peaks and greater variability in filter banks below 14, indicating dynamic frequency content. In contrast, Figure 5b displays the signal from samples of older age group with a more uniform distribution below filter 14, suggesting a stable spectral profile linked to reduced arterial flow variability. The signal from samples of younger age group exhibit greater temporal and spectral fluctuations, while the signal from samples of older age group are stable in lower-frequency bands, reflecting age-related changes in blood flow and arterial structure. The MFCCs capture the spectral envelope, with lower coefficients (1–5) showing concentrated energy in lower frequencies [11]. The signal from samples of younger age group also show more variability in middle MFCCs (6–11), while the signal from samples of older age group maintain a stable spectral pattern, aligning with their uniform filter bank output.



a. 24 years old b. 48 years old

Figure 5: Sample Spectrogram of MFCC Coefficients

3.3. MFCC across three different age groups

The signal from samples of younger age group exhibits greater variability across MFCC coefficients over time, especially in middle coefficients (5–11), indicating dynamic spectral changes, while the signal from samples of older age group displays stability in its coefficients pattern. This aligns with the more uniform signal observed in the older age filter bank outputs.

DOI: 10.54254/2755-2721/100/2025.18285

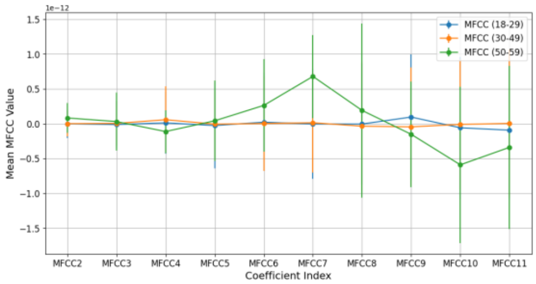


Figure 6: MFCC across different age groups

3.4. Principal component analysis with age groups

Principal Component Analysis (PCA) revealed distinct clustering patterns among age groups based on extracted features.

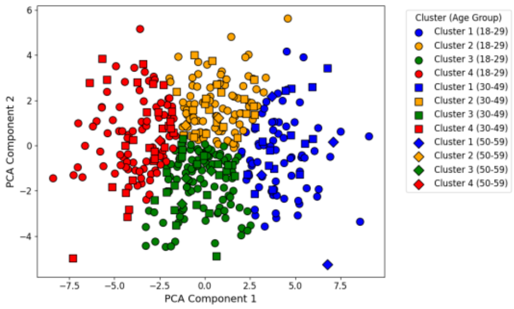


Figure 7: PCA plot with Cluster & age groups

As illustrated in Figure 7, four clusters (1 to 4) emerged corresponding to age groups: 18-29, 30-49, and 50-59. A clear separation was observed between younger (18-29) and older (50-59) groups along the first principal component (PC1), suggesting significant differences in their feature profiles

3.5. Experimental result with dataset

Experiments are carried out in this section to evaluate the performance of the proposed carotid artery signal classification algorithm. The first experiment tests the parameter sensitivity of the proposed RNN based classification algorithm. The training and validation accuracy plots, as shown in Figure 10, indicate that the model steadily improved during the early epochs, with validation accuracy stabilizing after approximately 50 epochs.

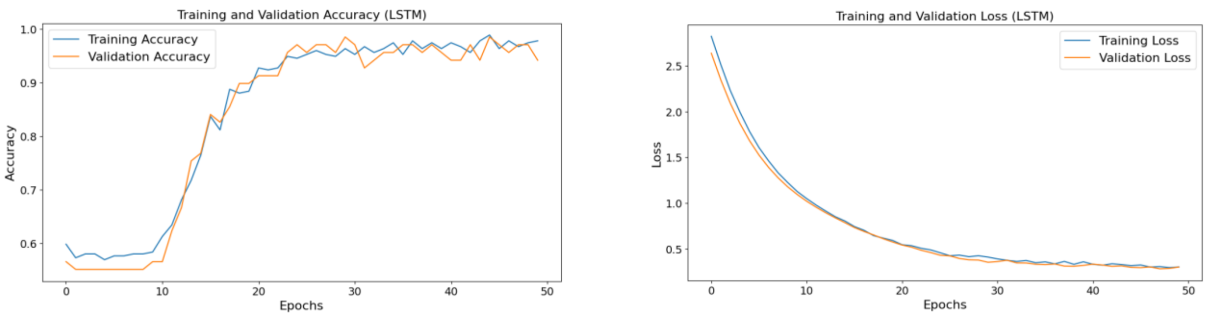


Figure 8: Accuracy & Loss Plot

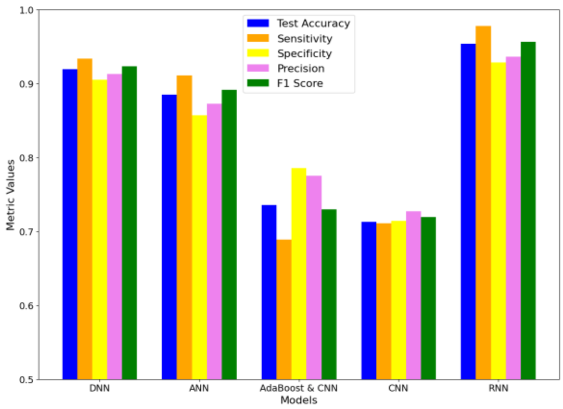


Figure 9: Comparison with SoA

The loss plot shows a rapid decrease in both training and validation loss within the first 20 epochs, after which both curves flatten, indicating that the model has effectively learned the data representations and is not overfitting significantly in terms of loss.

Figure 9 illustrates a comparison of different classification models (DNN, ANN, AdaBoost & CNN, CNN, and RNN) across various performance metrics: Test Accuracy, Sensitivity, Specificity, Precision, and F1 Score. The RNN model consistently outperforms other models, particularly in Test Accuracy and Sensitivity, indicating its strong ability to capture sequential dependencies in the data. While AdaBoost & CNN and CNN models exhibit lower performance across metrics, the DNN and ANN models also achieve competitive results, demonstrating effective generalization and balanced performance across all measured metrics

4. Discussion

The study examines age-related variations in Doppler ultrasound audio signals from the carotid artery using MFCCs and filter bank outputs. The signal from samples of younger age group showed greater variability, indicating more dynamic frequency content and arterial elasticity, while the signal from samples of older age group had a uniform spectral profile indicating stiffer arteries and less variability in blood flow. MFCC analysis revealed that the younger group experienced more dynamic spectral changes over time, particularly in middle MFCC coefficients, while the older group displayed a stable pattern. PCA visualization showed distinct clustering patterns across age groups, effectively capturing age-related variations. The RNN-based classification model demonstrated strong performance, but there is potential for refinement. Future enhancements, such as additional feature incorporation or model optimization, could further enhance the model's ability to capture age-related signal variations

5. Conclusion

The study demonstrates that MFCCs and filter bank outputs effectively capture age-related differences in Doppler ultrasound signals from the carotid artery. Signals are primarily concentrated in lower frequency bands, with younger individuals having greater spectral variability and older individuals having a stable profile. The RNN-based classification model showed promising results, but further refinement is needed for improved accuracy and robustness.

Acknowledgments

We would like to express our gratitude to Ms. Aswathy and Mr. Nijas, research scholars from the Medical Optics & Sensor Lab, Department of Biomedical Engineering, IIT Hyderabad, for their invaluable assistance in data collection. We also extend our appreciation to the Foundation for Center for Healthcare Entrepreneurship (CfHE) for their support in facilitating this research.

References

[1] *Dabbaghchian, S., Ghaemmaghami, M.P., & Esmailpour, M. (2010). Feature extraction using MFCC. Proceedings of the 2010 International Conference on Multimedia Computing and Information Technology (MCIT), 17–20.*

[2] *Huynh, H.T., & Ha, Q.M. (2020). Voice disorder classification based on MFCC and SVM. Journal of Computer Science and Cybernetics, 36(3), 201–210.*

[3] *Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., ... Nahavandi, S. (2021). A review of deep learning techniques for medical image segmentation. Artificial Intelligence in Medicine, 114, 102066. https://doi.org/10.1016/j.artmed.2021.102066*

[4] *Zhang, Z., Cui, P., & Liu, W. (2020). Deep learning for healthcare: Review, opportunities, and challenges. Briefings in Bioinformatics, 22(2), 805–818. https://doi.org/10.1093/bib/bbz092*

[5] *Suhn, T., et al. (2019). Design of an auscultation system for phonoangiography and monitoring of carotid artery diseases. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), 1776–1779.*

[6] *Iverson, B.L., & Dervan, P.B. (2006). The cervical bruit: sound spectral analysis related to severity of carotid arterial disease. Clinical Cardiology, 29(10), 462–465.*

[7] *Boussaa, M., Atouf, I., Atibi, M., & Bennis, A. (2016). ECG signals classification using MFCC coefficients and ANN classifier. Proceedings of the 2nd International Conference on Electrical and Information Technologies (ICEIT).*

[8] *Farzam, B., & Shirazi, J. (2014). The diagnosis of heart diseases based on PCG signals using MFCC coefficients and SVM classifier. IJISET - International Journal of Innovative Science, Engineering & Technology, 1(10).*

[9] *Deng, M., Meng, T., Cao, J., Wang, S., Zhang, J., & Fan, H. (2020). Heart sound classification based on improved MFCC features and convolutional recurrent neural networks. Neural Networks, 130, 22–32.*

[10] *Mu, X., & Min, C.H. (2023). MFCC as features for speaker classification using machine learning. Proceedings of the IEEE Conference on Machine Learning and Applications, 79-8-3503-3761-7/23/$31.00 ©2023 IEEE.*

[11] *Abbas, S.N., Abo-Zahhad, M., Ahmed, S.M., & Farrag, M. (2016). Heart-ID: human identity recognition using heart sounds based on modifying mel-frequency cepstral features. IET Biometrics, 5(4), 284–296. https://doi.org/10.1049/iet-bmt.2015.0033*