2024 CMBEC46
Toronto, ON
May 28-30, 2024
CMBES/SCGB

Fractional Fourier Time-Frequency Representation for Heart Sound Classification

Ebrahim A. Nehary[1] and Sreeraman Rajan[1]

[1] Department of System and Computer Engineering, Carleton University, Ottawa, Canada

Abstract— **Early detection of heart sounds can significantly reduce mortality rates by allowing physicians to intervene on time. However, manual heart sound analysis is subjective and relies heavily on the skills and experience of the physician. Fortunately, deep learning has emerged as a promising method for heart sound classification. Time-frequency representations (TFR) such as spectrograms, continuous wavelet transforms (CWT), and Mel-Frequency cepstral coefficients (MFCC) have been widely accepted input representations for heart sound representation. This study proposes a combination of fractional Fourier time-frequency representation (FrFT_TFR) and a deep learning model for heart sound classification. It uses a public dataset to demonstrate the efficacy of the proposed representation.Classification using a deep learning model with FrFT_TFR as input outperforms that obtained with spectrograms and MFCC as inputs by approximately 4% and by 13% over CWT as inputs. The results underscore the effectiveness of using FrFT_TFR for heart sound classification for initial heart sound diagnosis.**

Keywords— **Heart sounds, Classification, Fractional Fourier transform, Deep learning, MFCC**

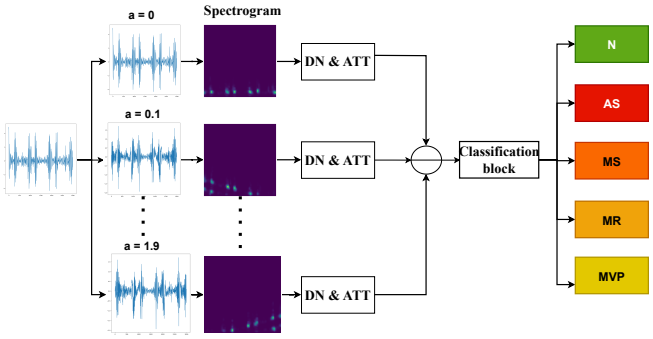
I. Introduction

Cardiovascular disease is a significant health threat with high mortality rates. In 2016, around 17.9 million deaths were recorded [1]. Early diagnosis of heart conditions can reduce the mortality rate. Heart sound auscultation is a non-invasive and cost-effective tool that can be used to detect cardiovascular diseases and, in particular, the functioning of heart valves. In particular, this cost-effective diagnosis is suitable for developing countries [2]. However, the accuracy of diagnosis relies on the skill and experience of the physician. An automated system can eliminate this shortcoming and aid physicians to diagnose cardiovascular diseases in their early stages.

Recent advancements in the field have demonstrated the potential of heart sound classification using deep learning, particularly with time-frequency representation (TFR) as input and hold promise for improving the efficiency and accuracy of cardiovascular disease diagnosis. Various TFRs such as Mel-frequency cepstral coefficients (MFCCs), continuous wavelet transform (CWT), and spectrogram have been proposed for heart sound classification. MFCCs have been widely used to transform the PCG signal from a one-dimensional time domain to a two-dimensional time-frequency domain. In [3], MFCC and PCG signals were used as inputs for 2-D and 1-D convolutional neural networks. MFCCs with time and time-frequency features were used as input to feed-forward neural networks in [4]. In [5], MFCCs with a spectrogram and mel spectrogram were used to train three independent VGG16 while MFCC and the discrete wavelet transform (DWT) were employed as input to the LSTM classifier in [6]. Also, MFCCs and time and frequency features were extracted from segmented signals and fed various classifiers such as LSTM and SVM in [7] for heart sound classification.

The spectrogram and CWT are commonly used TFRs in heart sound classification. Spectrogram of heart sound was used as input to ResNet18 model [8], custom CNN [2] for extracting abstract features to train a fully connected neural network for classification. In [9], spectrogram was used as inputs to the pre-trained VGG16, VGG19, and AlexNet models to extract features to train a support vector machine (SVM) classifier. CWT generated using the Morse wavelet was employed as input to pre-trained VGG16 and AlexNet for feature extraction and SVM was used for classification [10, 11], while AlexNet was used for both feature extraction and classification [12]. Morelet wavelet was employed as input to a shallow CNN [13].

Although various TFRs yield satisfactory results, this study aims to further enhance the classification performance by employing fractional Fourier Transform-based TFR (FrFT_TFR). This approach involves transforming the input signal into the fractional Fourier domain and subsequently computing TFR such as FrFT_spectrogram, FrFT_CWT, and FrFT_MFCC. The fractional Fourier Transform (FrFT) is a generalized form of the Fourier transform, symbolizing a rotation of a signal in a time-frequency plane. By accommodating signals in any intermediate domain between time and frequency, the FrFT can handle non-stationarity in signals and filter out noise with simultaneous temporal and spectral overlap with signals [14, 15]. The contributions of this paper are as follows: FrFT_TFR, specifically FrFT_spectrogram, FrFT_CWT, and FrFT_MFCC, are proposed to improve heart sound classification. A comparison between FrFT_TFR and conventional TFR is presented to demonstrate the effectiveness of FrFT_TFR for heart sound classification. Addition-



(a)

Fig. 1: Overall diagram of the proposed method. DN & ATT indicate DenseNet and attention.

ally, a deep learning model comprising DenseNet, attention mechanisms, and classification blocks is also proposed for classification. This is the first-ever study using FrFR_TFR for heart sound classification.

II. Methodology

The overall proposed diagram is shown in Figure 1. The proposed method transforms the heart sound signal into the fractional Fourier transform domain, followed by calculating TFRs of the transformed signals, such as spectrogram, CWT, and MFCC. Subsequently, the DenseNet block is employed to extract features from the TFRs, and these features are then fed into an attention method to recalibrate the extracted information. Finally, the attention features are concatenated and input into the classification block to yield the final classification results.

A. Fractional Fourier transform

The fractional Fourier Transform (FrFT) stands as a generalized form of the Fourier transform, symbolizing a rotation of a signal in a time-frequency plane, and it enables a more flexible time-frequency representation. By accommodating signals in any intermediate domain between time and frequency, the FrFT proves advantageous for non-stationary signal processing such as heart sound [15, 16, 17]. The elements of the FrFT matrix F^{α} is given as

$$F^{\alpha}(m,n) = \sum_{k=0}^{N-1} u_k(m) \exp\left(\frac{-j\pi k\alpha}{n}\right) u_k(n), \qquad (1)$$

where $0 < m,n < N-1$, u is a discrete Hermite-Gaussian function of order 2, and $0 \leq \alpha \leq 2$ is a fractional order. Then, the FrFT transform is the matrix-vector multiplication of the transform matrix $F^{\alpha}(m,n)$ (as given in equation 1) with the signal vector. It is worth mentioning that when $\alpha = 0$ or $\alpha = 2$, $F^{\alpha}(m,n)$ is the identity matrix, and the transform is equivalent to the identity operator on the signal. Additionally, the Fourier transform is the spatial case of FrFT when $\alpha = 1$. The heart sound signal of s of length N is transformed into a corresponding real-valued signal in the fractional Fourier domain for each α in the following manner.

$$x^{\alpha} = \Re\{F^{\alpha} \times s\} \qquad (2)$$

where $\Re$ denotes the real part of the complex vector of length N.

B. Time-Frequency representations (TFRs)

For completeness, a summary of different Time-Frequency transforms (spectrogram, CWT, MFCC, and MFSC) is given in this section. The spectrogram is the squared magnitude of the short-Fourier transform (STFT) coefficients of the signal x^{α}. The STFT of signal x^{α} is given below

$$STFT(n,k) = \sum_{m=n-(N_w-1)}^{n} w(n-m)x^{\alpha}(m)\exp(\frac{-j2\pi mk}{N}), \qquad (3)$$

where w is window function with nonzero values of n from 0 to $N_w - 1$. The CWT of signal $x^{\alpha}(i)$ is described as [18]

$$CWT(a,b) = \frac{1}{|a|}\int_{-\infty}^{\infty} x^{\alpha}(t)\psi * \left(\frac{t-b}{a}\right) \qquad (4)$$

where ψ^* denotes the complex conjugate of the mother wavelet (Morse wavelet in this work). Where a and b are dilation and translation parameters, respectively. The Mel frequency cepstral coefficients (MFCCs) is another time-frequency representation of the signal [19]. The input signal undergoes pre-emphasis using the filter $H(z) = (1-0.9z^{-1})$, followed by framing into short segments. The Fourier transform of these short frames is computed, and the power of the transformed frames is then mapped into the mel-scale using triangular overlapping windows. The mel-scale is defined by the equation

$$\text{Mel}(f) = 2595\log_{10}\left(1+\frac{f}{700}\right), \qquad (5)$$

where f represents the frequency in Hz. Subsequently, the logarithm of the discrete cosine transform (DCT) of the mel-scaled map is calculated to derive the MFCC features.

C. Deep learning models

The deep learning model comprises three blocks: the DenseNet block, the attention block, and the classification block. The DenseNet block is employed to extract features from the input TFRs and comprises a dense and transition modules. The dense module consists of four convolutional layers with the ReLU activation function, followed by batch normalization and dropout layers with a ratio of 0.5 to prevent

overfitting. The filter size is set to 3×3, the number of filters is 32, and the compression is 0.5. The output of the convolutional layers is then fed to the transition module, which includes a convolutional layer with a filter size of 1×1, followed by dropout with a ratio of 0.5 and average pooling layers with pooling size of 2×2 to reduce the number of model parameters.

The attention block is inspired by [2], and it contains two submodules: the channel attention module (CHA) and the coordinate attention module (COA). The CHA recalibrates channel-wise features using global average pooling to obtain channel-wise statistics. The CHA is defined as $CHA = \mathrm{ReLU}\left(W_2\sigma(W_1\mathrm{GAP}(F))\right)$, where W_1 and W_2 are the weights of fully connected layers, F is the feature, GAP is the global average pooling layer, and σ is the sigmoid function. On the other hand, the COA is employed to capture both channel relationships and long-range dependencies. It first encodes vertical and horizontal pooling along the respective coordinates, concatenates the encoded features, and feeds them to a convolutional layer with an H-Swish activation function. The resulting output is split into separate tensors along the spatial dimension. Each split tensor is then passed through a convolutional layer and a sigmoid activation function to obtain two scores (s_1 and s_2). Finally, the attention score is computed by multiplying both scores ($s_1 * s_2$). Notably, the convolutional layer in COA has a filter size of 1.

The classification block comprises a batch normalization layer, an average pooling layer, a fully connected layer with 5 neurons, and a softmax activation function.

D. Data preparation and training model

The proposed method is assessed using the dataset provided in [20]. This dataset comprises a total of 1000 records across five classes: 200 records for normal (N), 200 for aortic stenosis (AS), 200 for mitral stenosis (MS), 200 for mitral regurgitation (MR), and 200 for mitral valve prolapse (MVP). The dataset, sampled at a frequency of 8000 Hz, was collected from diverse sources, including two books and 48 websites. A five-fold cross-validation is applied to partition the dataset into training and testing sets, with the results reported as the mean over five iterations. The evaluation metrics are sensitivity, specificity, precision, f1 score, and accuracy.

The dataset is pre-processed by resampling each record to 1000 Hz and filtering it using a Butterworth bandpass filter of order 6 with a cut-off frequency between 25 and 400 Hz. This filtering process is employed to remove baseline drift and interference noise, such as noise from blood flow, environmental factors, and physical contact between the body and the sensor [21]. A constant window with a length of 3 seconds is extracted from each record, and if the recording is less than 3 seconds, zero-padding is applied to achieve a consistent length. Subsequently, the FrFT is employed with a fractional order ranging from 0 to 1.9 in steps of 0.1. In other words, each record is represented in 20 different fractional Fourier domain. Following this, the time-frequency representation (spectrogram, CWT, and MFCC) of each transformed signal is computed. For the spectrogram, a hamming window with a length of 125 ms and 64 ms overlapping is used, and the same window length and overlapping are employed for both MFCC with 20 bands while the Morse wavelet has $a = 10$ and $b = 60$. The time-frequency representation is then fed into the DenseNet block to extract features, as shown in Figure 1. The extracted features are used as input to the attention block to emphasize useful features and reduce irrelevant ones. Finally, the recalibrated features from each fractional time-frequency domain are concatenated and fed into the classification block for the final classification into five classes. The deep learning model is trained using categorical cross-entropy as the loss function, Adam as the optimization method, and 300 epochs. A model checkpoint is implemented to save the best model based on the evaluation loss, and the learning rate is decreased by a factor of 10 for every 50 epoch, with the initial learning rate set at 0.0001.

Table 1: Comparison performance of time-frequency representation with/without FrFT transform.

TFR	Sensitivity	Specificity	Precision	F1 score	Accuracy
FrFT_Spectrogram	95.51	98.85	95.85	95.37	95.40
Spectrogram	90.46	97.75	92.45	89.98	91.00
FrFT_CWT	94.35	98.60	94.59	94.33	94.40
CWT	81.57	95.32	84.36	80.70	81.20
FrFT_MFCC	**98.42**	**99.60**	**98.42**	**98.39**	**98.40**
MFCC	94.52	98.60	94.65	94.43	94.40

III. Results

Table 1 shows the results obtained using FrFT_TFR (FrFT_Spectrogram, FrFT_CWT, and FrFT_MFCC) in comparison to using TFRs only (Spectrogram, CWT, and MFCC). Table 1 demonstrates that FrFT significantly improves the classification performance. Specifically, FrFT_Spectrogram outperforms spectrogram by approximately 4%, FrFT_CWT surpasses CWT by approximately 13%, and FrFT_MFCC exceeds MFCC by approximately 4% based on accuracy. Overall, FrFT consistently provides superior results for heart sound classification. The most significant improvement in results is observed with FrFT_MFCC, highlighted in bold in Table 1.

These findings emphasize the effectiveness of integrating FrFT transformations with Spectrogram, CWT, or MFCC in the proposed deep learning model. Furthermore, Figure 2 shows the confusion matrices for the proposed methods, comparing the use of FrFT against not using FrFT for each class. Once again, the confusion matrix with FrFT demonstrates superior performance compared to the time-frequency repre-

sentation without FrFT. Notably, the best results are achieved when utilizing FrFT_MFCC as input to the proposed method, with correct classifications of 198, 192, 198, 198, and 198 for N, MVP, MS, MR, and AS classes, respectively (see Figure 2 (c)). In contrast, without FrFT, the correct classifications decrease, especially for the normal (N) class, dropping from 198 to 177 (see Figure 2 (f)). This decline in correctly classified records, particularly for the normal class, highlights the significance of employing the FrFT transform in the proposed method, as supported by the confusion matrices.

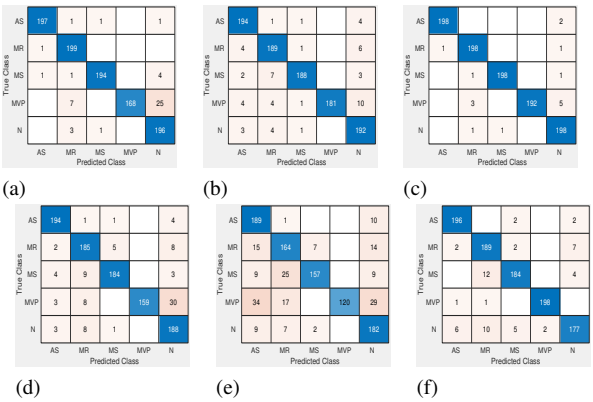


Fig. 2: The confusion matrices depict the performance of the proposed methods with and without FrFT transforms in first and second rows respectively. Spectrogram, CWT and MFCC results are shown in the first, second and third columns respectively.

IV. Conclusion

A deep learning-based heart sound classification model using a FrFT_TFR for transforming the heart sound signal was proposed. FrFT_TFR produces better results than traditional TFR, such as spectrogram, CWT, and MFCC. However, the improvement is achieved at the expense of the number of parameters in the classifier model. The classifier model using traditional TFR has fewer parameters ($40K$) than the models that use FRFT_TFR ($700K$). These findings underscore the effectiveness of using FrFT_TFR as input to a deep learning model to enhance classification performance. Future work will focus on implementing advanced techniques to determine an optimal fractional order to further improve classification performance and reduce the number of parameters in the proposed models.

Conflict of Interest

The authors declare that they have no conflict of interest.

ACKNOWLEDGEMENTS

This work was financially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

1. WHO, “Cardiovascular diseases, cvd fact sheet, world health organisation,” Accessed January 21, 2024 [Online]. [Online]. Available: ”www.who.int/health-topics/cardiovascular-diseases#tab=tab_1”
2. J. Chen *et al.*, “A robust deep learning framework based on spectrograms for heart sound classification,” *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 2023.
3. F. Noman, C.-M. Ting, S.-H. Salleh, and H. Ombao, “Short-segment heart sound classification using an ensemble of deep convolutional neural networks,” in *IEEE ICASSP*, 2019, pp. 1318–1322.
4. M. Abdollahpur, A. Ghaffari, S. Ghiasi, and M. J. Mollakazemi, “Detection of pathological heart sounds,” *Physiol. Meas.*, vol. 38, no. 8, p. 1616, 2017.
5. J. Wu *et al.*, “Applying an ensemble convolutional neural network with Savitzky Golay filter to construct a phonocardiogram prediction model,” *Appl. Soft Comput.*, vol. 78, pp. 29–40, 2019.
6. B. Ahmad *et al.*, “Automatic classification of heart sounds using long short-term memory,” in *IEEE ICOSST*, 2021, pp. 1–6.
7. F. Khan, A. Abid, and M. Khan, “Automatic heart sound classification from segmented/unsegmented phonocardiogram signals using time and frequency features,” *Physiol. Meas.*, vol. 41, no. 5, p. 055006, 2020.
8. S. Takezaki and K. Kishida, “Construction of CNNs for abnormal heart sound detection using data augmentation,” in *Proc. IMECS*, 2021, pp. 1–6.
9. F. Demir, A. Şengür, V. Bajaj, and K. Polat, “Towards the classification of heart sounds based on convolutional deep neural network,” *Health Inf. Sci. Syst.*, vol. 7, pp. 1–9, 2019.
10. Z. Ren, N. Cummins, V. Pandit, J. Han, K. Qian, and B. Schuller, “Learning image-based representations for heart sound classification,” in *ICDHT*, 2018, pp. 143–147.
11. H. Alaskar *et al.*, “The implementation of pretrained AlexNet on PCG classification,” in *ICIC*, 2019, pp. 784–794.
12. S. Singh, S. Majumder, and M. Mishra, “Classification of short unsegmented heart sound based on deep learning,” in *IEEE I2MTC*, 2019, pp. 1–6.
13. A. Bhardwaj, S. Singh, and D. Joshi, “Explainable deep convolutional neural network for valvular heart diseases classification using PCG signals,” *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–15, 2023.
14. C. Candan *et al.*, “The discrete fractional fourier transform,” *IEEE Trans. Signal Process*, vol. 48, no. 5, pp. 1329–1337, 2000.
15. A. McBride and F. Kerr, “On Namias’s fractional Fourier transforms,” *IMA J. Appl. Math.*, vol. 39, no. 2, pp. 159–175, 1987.
16. L. B. Almeida, “The fractional Fourier transform and time-frequency representations,” *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 3084–3091, 1994.
17. E. Sejdić, I. Djurović, and L. Stanković, “Fractional Fourier transform as a signal processing tool: An overview of recent developments,” *Signal Process.*, vol. 91, no. 6, pp. 1351–1369, 2011.
18. S. Olhede and A. Walden, “Generalized morse wavelets,” *IEEE Trans. on Signal Process.*, vol. 50, no. 11, pp. 2661–2670, 2002.
19. Z. Abduh, E. Nehary, M. Abdelwahed, and Y. Kadah, “Classification of heart sounds using fractional Fourier transform based mel-frequency spectral coefficients and traditional classifiers,” *Biomed. Signal Process. Control*, vol. 57, p. 101788, 2020.
20. Yaseen, G.-Y. Son, and S. Kwon, “Classification of heart sound signal using multiple features,” *Applied Sciences*, vol. 8, no. 12, pp. 2344–2358, 2018.
21. S. Das, D. Jyotishi, and S. Dandapat, “Automated detection of heart valve diseases using stationary wavelet transform and attention based hierarchical LSTM network,” *IEEE Trans. Instrum. Meas.*, 2023.

Author: Ebrahim A. Nehary
Institute: Carleton University
Street: 1125 Colonel By Dr
City: Ottawa
Country: Canada
Email: ebrahimali@cmail.carleton.ca