Available *online* at : http://jnte.ft.unand.ac.id/

Jurnal Nasional Teknik Elektro

| ISSN (Print) 2302-2949 | ISSN (Online) 2407-7267 |

Enhanced Identification of Valvular Heart Diseases through Selective Phonocardiogram Features Driven by Convolutional Neural Networks (SFD-CNN)

Muhammad Rafli Ramadhan[1,2], Satria Mandala[1,2], Rafi Ullah[3], Wael M.S. Yafooz[4], Muhammad Qomaruddin[5]

[1] *Department of Informatics, School of Computing, Telkom University, Bandung 40257, Indonesia*
[2] *Human Centric (HUMIC) Engineering, Telkom University, Bandung 40257, Indonesia*
[3] *Computer and Information Sciences Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar Perak, 32610, Malaysia*
[4] *Computer Science Department, Taibah University, Madinah, 42353, Saudi Arabia*
[5] *Informatics Engineering, Faculty of Industrial Technology, Universitas Islam Sultan Agung, Semarang, 50112, Indonesia*

ARTICLE INFORMATION

Received: January 11, 2024
Revised: January 20, 2024
Accepted: March 03, 2024
Available online: March 29, 2024

KEYWORDS

Valvular heart disease, PCG, classification, deep learning

CORRESPONDENCE

Phone: +62 821 2048 1404
E-mail: satriamandala@telkomuniversity.ac.id

A B S T R A C T

Valvular Heart Disease (VHD) is a significant cause of mortality worldwide. Although extensive research has been conducted to address this issue, practical implementation of existing VHD detection results in medicine still falls short of optimal performance. Recent investigations into machine learning for VHD detection have achieved commendable accuracy, sensitivity, and robustness. To address this limitation, our research proposes utilizing Selective Phonocardiogram Features Driven by Convolutional Neural Networks (SFD-CNN) to enhance VHD detection. Notably, SFD-CNN operates on phonocardiogram (PCG) signals, distinguishing itself from existing methods based on electrocardiogram (ECG) signals. We present two experimental scenarios to assess the performance of SFD-CNN: one under default parameter conditions and another with hyperparameter tuning. The experimental results demonstrate that SFD-CNN surpasses other existing models, achieving outstanding accuracy (96.80%), precision (93.25%), sensitivity (91.99%), specificity (98.00%), and F1-score (92.09%). The outstanding performance of SFD-CNN in VHD detection suggests that it holds great promise for practical use in various medical applications. Its potential lies in its ability to accurately identify and classify VHD, enabling early detection and timely intervention. SFD-CNN could significantly improve patient outcomes and reduce the burden on healthcare systems. With further development and refinement, SFD-CNN has the potential to revolutionize the field of VHD detection and become an indispensable tool for healthcare professionals.

INTRODUCTION

Valvular Heart Disease (VHD) is caused by functional disorders in one or more heart valves, such as the tricuspid, pulmonary, mitral, and aortic valves. VHD can result from congenital or birth defects or infections that damage the heart valves [1]. The presence of VHD was found to be independently associated with a 65% increased risk of all-cause mortality and a 72% increased risk of cardiovascular mortality [2]. While electrocardiogram (ECG) technology can identify VHD, it is expensive. Therefore, with the growing development of technologies such as healthcare applications [3], a more cost-effective approach is needed to detect VHD using phonocardiogram (PCG) signals. However, identifying heart sound abnormalities indicating VHD through PCG has limitations, as it requires medical experts with skill and experience. Furthermore, this procedure may be subject to inaccuracies depending on the physician's hearing health. Therefore, some researchers suggest using deep learning algorithms to interpret patient phonocardiogram signals [4], [5]. In addition to deep learning, other researchers also employ standard machine learning algorithms to identify VHDs, as demonstrated in [6], [7], [8], [9], [10]. Another method researchers use to identify VHD is ensemble learning (EL). Several researchers, such as in [11], [12], [13] utilized the method in phonocardiogram signals.

Detecting valvular heart disease (VHD) based on phonocardiograms involves several sequential steps. Commencing with preprocessing, this initial and pivotal step aims to eliminate noise and minimize abnormal data [11], [14]. Following preprocessing, the feature extraction phase identifies and extracts crucial elements from the phonocardiogram signal relevant to VHD. This involves activities such as recognizing frequency patterns or sound waveforms. The extracted features are then input into a classifier, typically artificial neural networks [5], [15] to determine the presence of VHD by analyzing the signal's waveforms.

Numerous studies have explored artificial intelligence-based VHD detection on phonocardiograms. Generally, these studies

https://doi.org/10.25077/jnte.v13n1.1184.2024

[7], [8], [16], [17], [18] employ signal morphology-based feature extraction, such as Discrete Wavelet Transform (DWT) and Mel Frequency Cepstral Coefficients (MFCC), combined with classic machine learning algorithms like Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Multi-Layer Perceptron (MLP). Some researchers opt for different classification algorithms, such as deep learning (DL) [16], [19] or ensemble learning (EL) Sinha Roy et al. [20]. However, suboptimal results often occur due to incorrect feature extraction algorithms and classifier configurations.

In studies conducted by [7], [8], [17], [18], the accuracy obtained was consistently less than 95%. For instance, Yaumil et al. [7] applied MFCC with an SVM classifier, while Bhole et al. [8] incorporated MFCC, Zero Crossing Rate (ZCR), peak amplitude, and classifiers such as KNN, Adaboost, and SVM. Li et al. [17] experimented with SVM and Twin Support Vector Machine (TWSVM) classifier configurations using features from wavelets and entropy. Binta et al. [18] explored the combination of features from MFCC and DWT with an SVM classifier.

Research adopting DL [16], [19] as a classifier demonstrated high VHD detection accuracy, but concerns about overfitting arose. For instance, Alqudah et al. [16] combined an FFT-based feature extraction algorithm with DL classifiers like CNN, CNN-SVM, and CNN-KKN. Meanwhile, Karhade et al. [19] used time-frequency-domain-based feature extraction with a CNN classifier.

Similarly, researchers employ the EL algorithm [12], [13], [20] as a classifier. For instance, Sinha Roy et al. [20] encountered accuracy challenges, achieving only around 96%. Sinha Roy et al. [20] also employed feature extraction algorithms such as Root Mean Square (RMS), signal energy, signal power, ZCR, skewness, kurtosis, and DWT. For the classifier, Sinha Roy et al. [20] considered the use of Random Forest (RF), Decision Tree (DT), Gradient Boosting (GB), and Xgboost (XGB) algorithms. While Ghosh et al. [12] used the RF classifier with time-frequency feature extraction, and Tuncer et al. [13] used their own proposed method, such as Phonocardiogram- Tent pooling (PCG-TEP) and Iterative neighborhood component analysis (INCA) used the DT classifier whose accuracy is still less than 95%.

To address the challenges identified in previous studies, this research proposes an improved approach to detecting VHD. Specifically, this study considers a DWT-based feature extraction algorithm with a CNN classifier. In contrast to earlier studies, this research stands out by meticulously selecting 100 combinations of Daubechies wavelet basis functions and level decompositions, ranging from one to ten. Furthermore, the setting parameters of the classifier used in this research also differ from previous studies. Instead of following established parameters, this study opts for default CNN settings combined with hyperparameter tuning using gridSearchCV. This distinctive approach aims to enhance the accuracy and robustness of valvular heart disease (VHD) detection, setting it apart from conventional methods employed in the field. This study contributes significantly to heart sound analysis, encapsulated in the proposed Selective Phonocardiogram Features Driven by Convolutional Neural Networks (SFD-CNN). A notable aspect of this research lies in its emphasis on augmenting accuracy rates through implementing a selective features approach to Phonocardiogram (PCG) signals, a pioneering step many researchers in the field still need to explore. The study systematically compares various machine learning models applied to the same dataset, comprehensively evaluating their effectiveness. Furthermore, the research meticulously assesses the SFD-CNN model through two distinct experimental scenarios: one without tuning and another with hyperparameter tuning.

RELATED WORK

This section delves into Valvular Heart Disease (VHD) research that utilizes machine learning techniques. Various studies, including those by researchers in [5], [15], [21], [22], have suggested employing CNN classification and DWT feature extraction methods to enhance VHD detection.
Alkhodari et al. [5] employed the CNN-BiLSTM approach along with DWT and CWT, achieving an accuracy of 87.31%, sensitivity of 92.78%, and specificity of 79.48%. Similarly, Roy et al. [21] utilized a CNN Residual Net model with MFCC and DWT, achieving notable results with 97.32% accuracy, 98.42% precision, 98.32% recall, and a 98.35% F1-score. Aljohani et al. [15] applied CNN classification with DWT and MFCC, obtaining a precision, recall, and F1-score of 95.5%. However, their reported results need to be more precise.

Flores et al. [22] utilized CNN classification with DWT, MFCC, and CWT, reporting validation results of 99.20% accuracy, 99.80% precision, 95.50% recall, and a 95.5% F1-score. Notably, they trained on 80% of the data without separate testing, thus needing more accuracy values for the test data. Roy et al. [21] used various CNN classification models, with the Xception CNN model achieving the highest accuracy rate of 99.43%. However, the drawback of their research is the extended training time required to achieve 99.43% accuracy.

In another study, Alqudah et al. [23] applied a CNN architecture to the AOCTNet system, achieving accuracy and recall rates of 98.70% and 97.10%, respectively. However, the research's weakness is the extended time required for PCG signal testing due to the image data generation method.
Karhade et al. [19] took a unique approach by converting PCG signals into images using the Frequency-Domain Polynomial Chirplet turn (FDPCT) and STFT methods. The CNN algorithm yielded a classification accuracy of 99.48%. However, during Leave-one-out cross-validation (LOOCV) with CNN classification, overfitting occurred with an accuracy of 100%.

Singh et al. [24] utilized the CNN algorithm with the CWT method and Butterworth filter to denoise PCG signals, achieving an accuracy of 87.96%. The potential for improvement lies in combining feature extraction methods and exploring additional deep learning techniques for more precise pattern identification.
Arslan et al. [25] employed the Hilbert-Huang transform (HHT) for feature extraction, comparing machine learning techniques (SVM, KNN) with deep learning techniques (DNN, MLP). DNN achieved an accuracy level of 98.00%, but the drawback is the longer training time associated with the HHT method.

METHODS

The methods section will contain a detailed and logical explanation of the process and sequence of method steps. The dataset was divided into training and testing datasets with a ratio of 90:10. The training dataset was used for five k-fold cross-validation. This method increases the robustness and reliability of the training process. To enhance VHD detection, 100 CNN models were developed using a combination of Daubechies (DB) wavelet basis function and decomposition level (DL) with default parameters. The performance of each model is evaluated using the training dataset as input to the K-fold cross-validation process. This research uses the K-fold cross-validation method to find the best fold on the training dataset. In addition, two experimental scenarios were performed on the CNN classification algorithm: one using default parameters and the other hyperparameter tuning. This approach aims to optimize the algorithm's performance and improve the accuracy of VHD detection on the dataset. The CNN model that performed best on the training data was selected to predict the testing dataset.

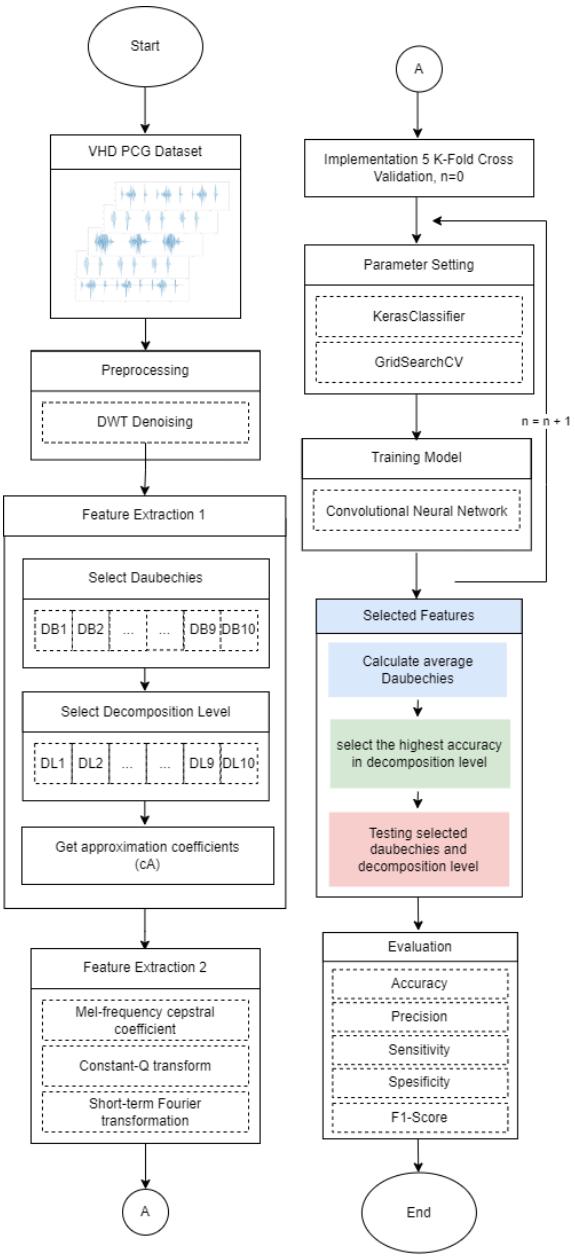


Figure 1. Research workflow

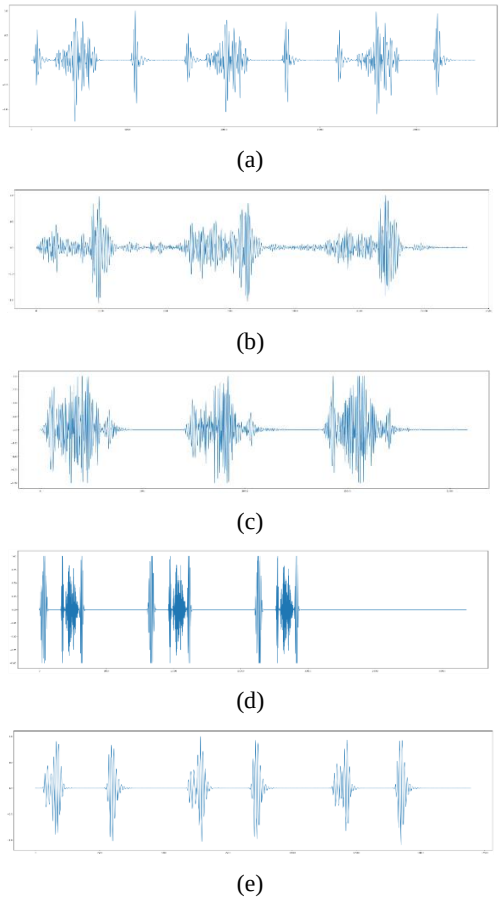
As shown in Figure 1, the selected features method in this research is performed by taking the best Daubechies wavelet function and level decomposition. These results are then evaluated using accuracy, precision, sensitivity, specificity, and F1-score metrics.

Dataset

Yaseen et al. dataset was utilized throughout the training phase [6]. Table 1 shows that the collection contains 1000 heart sound samples, with 200 samples per type divided into five categories. In Figure 2, the signal of the sample data for each VHD category is shown; there are five different types of valve prolapse: Mitral Valve Prolapse (MVP), Aortic Stenosis (AS), Mitral Regurgitation (MR), Mitral Stenosis (MS), and Normal (N). Figures 5, 6, and 7 demonstrate that the spectrogram and the dataset reveal different patterns for each data sample. This highlights the unique characteristics of each entity's frequency spectrum and provides a deeper understanding of the data set.

Table 1. Details of dataset

Binary Label	Multiclass Label	Number of Record
Normal	Normal (N)	200
Abnormal	Aortic Stenosis (AS)	200
	Mitral Regurgitation (MR)	200
	Mitral Valve Prolapse (MVP)	200
	Mitral Stenosis (MS)	200
Total		1000



(a)

(b)

(c)

(d)

(e)

Figure 2. Sample of dataset Yaseen et al.; (a) Mitral Stenosis; (b) Mitral Regurgitation; (c) Aortic Stenosis; (d) Mitral Valve Prolapse; (e) Normal

https://doi.org/10.25077/jnte.v13n1.1184.2024

Preprocessing

At this phase, the dataset heart sound data is processed before being given to the model. For instance, noise is removed using DWT, and heart sounds with irregular amplitudes are equalized in amplitude without impacting their quality.

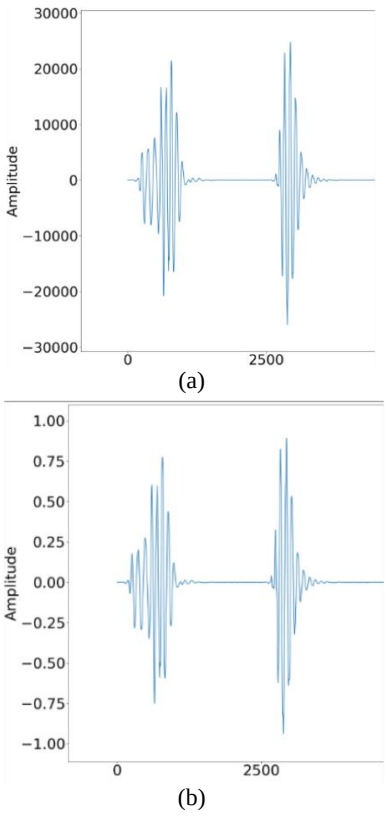


Figure 3. DWT Denoising. (a) Before denoising; (b) After denoising

Figure 3 shows that denoising reduces the noise in the PCG signal and normalizes the signal within the range of -1.0 to 1.0. The signal remains unaltered, mainly indicating that a significant decrease in amplitude does not affect the signal attributes. Maintaining a stable amplitude facilitates signal classification and recognition, thereby increasing pattern recognition accuracy.

Feature Extraction

DWT decomposition Level can break an audio stream into smaller bands [11]. In heart sound analysis, selecting the appropriate DWT level and figuring out the breakdown at the optimal level are crucial [26], [27].

$$W(a,b) = \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{a}}\psi * \left\{\frac{t-b}{a}\right\} dt \quad (1)$$

Equation 1 shows the basis function of the wavelet used in DWT. Where (t) is generated from a single mother wavelet by dilation and translation [26].

In this study, feature extraction 1 is detailed in Figure 4. Figure 4 shows that the chosen feature was DWT's approximation coefficient (cA), fed into MFCC, CQT, and STFT.

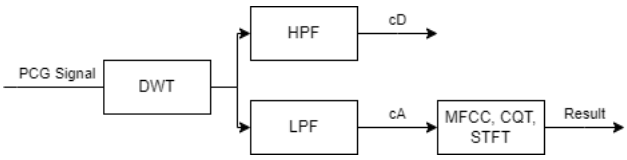


Figure 4 Feature Extraction 1 Flow

Mel-frequency cepstral coefficients (MFCC) frequently extract audio data into several parameters. The audio stream is first divided into many frames by MFCC, after which the amplitude spectrum is obtained using the logarithmic operation and the Discrete Fourier Transform (DFT). After doing some mel-scaling and filter bank changes, the MFCC is derived by applying a discrete cosine transform to the earlier findings [28]. Figure 5 shows the result of the MFCC spectrogram.

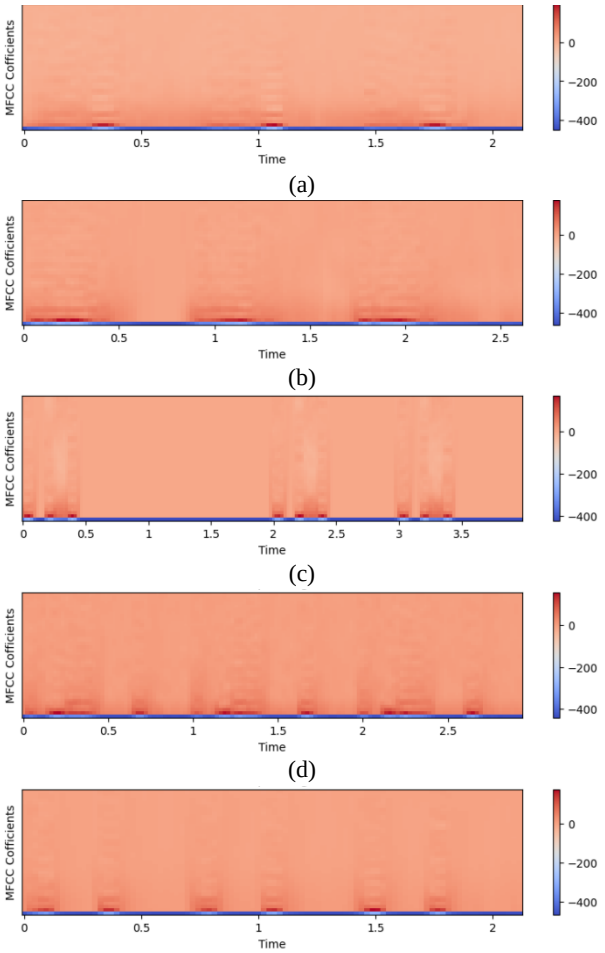


Figure 5. Spectrogram MFCC; (a) Mitral Regurgitation; (b) Aortic Stenosis; (c) Mitral Stenosis; (d) Mitral Valve Prolapse; (e) Normal

$$c[n] = \sum_{m=0}^{M-1} S[m]\cos\left(\frac{\pi n}{M}\left(m-\frac{1}{2}\right)\right), n = 0,1,2,\ldots,M \quad (2)$$

Equation 2 shows that the spectrum's discrete cosine transform (DCT) is calculated to obtain the MFCC. Where M is the number of filter banks in total, the remaining coefficients are utilized for training and testing, with the first being eliminated. [28]

Constant-Q transform (CQT) signal processing converts a collection of information into a frequency range. Particular

distinctions between the CQT and the Fourier transform include a fixed bandwidth filter and a logarithmic frequency scale. A practical approach can be used to compute the CQT, which is similar to a wavelet transform but has a higher frequency resolution [29]. Figure 6 shows the result of the CQT spectrogram.

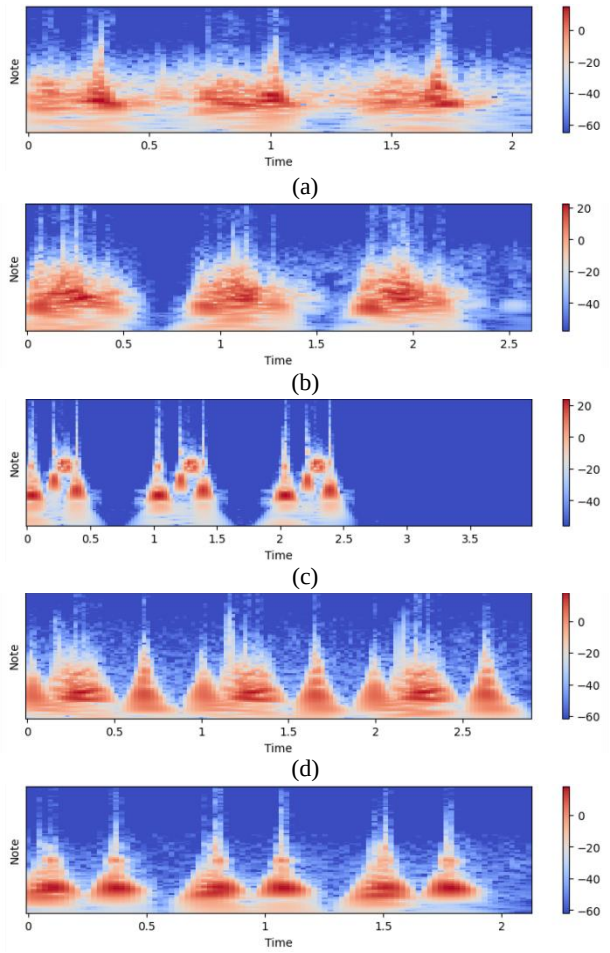
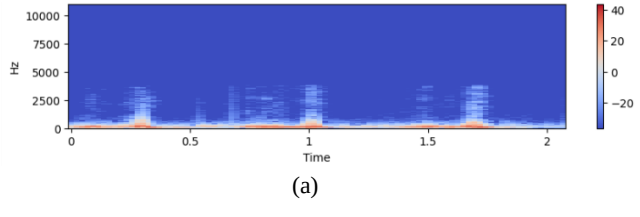


Figure 6. Spectrogram CQT; (a) Mitral Regurgitation; (b) Aortic Stenosis; (c) Mitral Stenosis; (d) Mitral Valve Prolapse; (e) Normal.

$$X^{CQT}(k,n) = \sum_{j=n-\lfloor N_k/2 \rfloor}^{n+\lfloor N_k/2 \rfloor} x(j) a_k^*(j - n + N_k/2) \quad (3)$$

Equation 3 shows the CQT formula. Where k is equal to 1, 2,... The CQT frequency bin is indicated by K, rounding to negative infinity is indicated by $\lfloor \cdot \rfloor$, and the complex conjugate of ak (n) is indicated by a $*$ k (n). A complex waveform is the foundation for the function ak (n) [29]

Short-term Fourier transformation (STFT) The twelve pitch classes used in audio studies are described by STFT in terms of intensity, which allows for distinguishing pitch class profiles of audio signals from one another. The pitch class profiles of audio signals can be distinguished using it [30]. Figure 7 shows the result of the SFFT spectrogram.



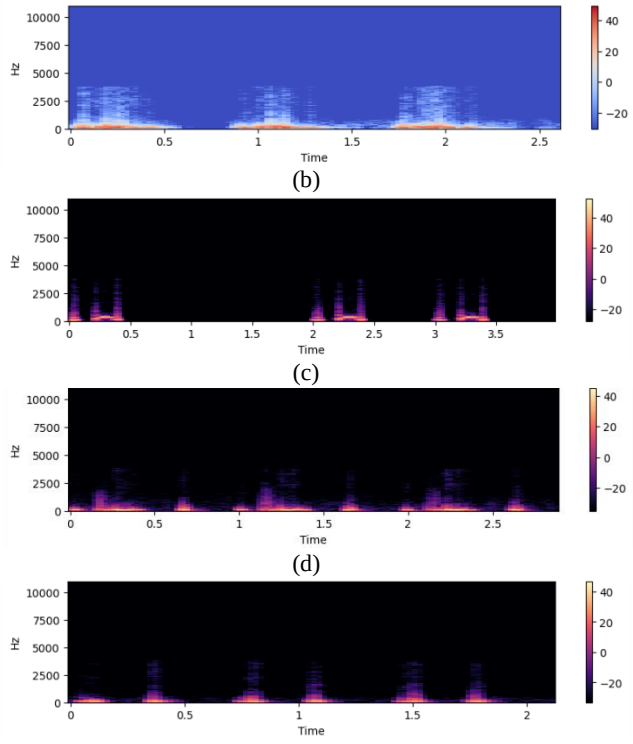


Figure 7. Spectrogram SFFT; (a) Mitral Regurgitation; (b) Aortic Stenosis; (c) Mitral Stenosis; (d) Mitral Valve Prolapse; (e) Normal.

$$X_{STFT}[k,n] = \sum_{m=0}^{N-1} x[m + nH] w[m] W_N^{km}, 0 \le k \le N - 1, 1 \le H \le N \quad (4)$$

Equation 4 shows the STFT formula. The window length is N, the time frame index is n, the frequency index is k, the input signal is x[m], and the window function is w[m]. In this case, the skip length is H, while the overlap length between consecutive frames is N − H. Moreover, (N − H)/N represents the overlap ratio between successive frames. The window function w[m] multiplies the signal x[m] at a specified time nH [31].

CNN Classifier

In this research, the classification algorithm used to train the model from the DWT, CQT, SFFT, and MFCC datasets is CNN. Table 2 details the CNN architecture used in this study. The tuning architecture uses the L2 regularizer in the fine-tuning model, distinguishing it from the non-tuning model. Softmax is used in the output layer of this study as it offers probability calculations for each class.

Table 2. CNN architecture layer

Layer	Activation Function	Regularizers
Conv1D(64)	ReLU	-
MaxPooling1D	-	-
Conv1D(32)	ReLU	-
Flatten	-	-
Dense(64)	ReLU	L2 (Fine Tuning)
Dense(5)	softmax	-

https://doi.org/10.25077/jnte.v13n1.1184.2024

Fine-Tuning

In this study, an automated method using GridSearchCV is required for parameter refinement and to improve the overall efficiency of model construction. Table 3 shows the hyperparameter configuration used in this study. The batch sizes are quite different between fine-tuning and nonfine tuning; nonfine tuning uses 64 batch sizes, and fine-tuning uses 16. The SFD-CNN fine-tuning model utilizes a smaller batch size compared to the non-tuning model. This is because a smaller batch size enhances the model's ability to adapt to variable data.

Table 3. Parameter used for model training

Parameter	Value
Epoch	50
Batch Size (Non Tuning)	64
Batch Size (Fine Tuning)	16
Optimizer	Adam
Activation Function	Softmax
Learning Rate	0.0001

Selected Features

The selected features are crucial for our research. To determine the optimal combination and level, follow these steps: First, calculate the average accuracy of the Daubechies wavelets and select the one with the highest average. Second, level decomposition is performed by choosing the Daubechies wavelet with the highest accuracy.

Metrics Evaluation

In this research, five metrics, accuracy, precision, sensitivity, specificity, and F1-score, are used to evaluate the performance of the proposed SFD-CNN. Equations 4 to 9 are the details of the metrics.

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (5)$$

$$precision = \frac{TP}{TP + FP} \quad (6)$$

$$specificity = \frac{TN}{TN + FP} \quad (7)$$

$$sensitivity = \frac{TP}{TP + FN} \quad (8)$$

$$f1score = 2 \times \frac{precision \times sensitivity}{sensitivity + precision} \quad (9)$$

Where TP and TN represent the total of correctly classified samples up to N, divided into VHD (True Positive) and non-VHD (True Negative). where N sample beats represent the sum of the non-VHD (False Negative) and VHD (False Positive) misclassifications [21], [32].

RESULTS AND DISCUSSION

This section describes the results and discusses the performance of the SFD-CNN model. Figure 8 shows that the amplitude increases at levels 1-5, but the decomposition of the PCG signal is not distorted. At levels 6-10, the PCG signal becomes more spartan, but the signal undergoes significant structural changes from the original signal. Since the PCG signal decomposition becomes similar to all signals at levels 8-10, the accuracy at levels 8-10 is relatively poor.

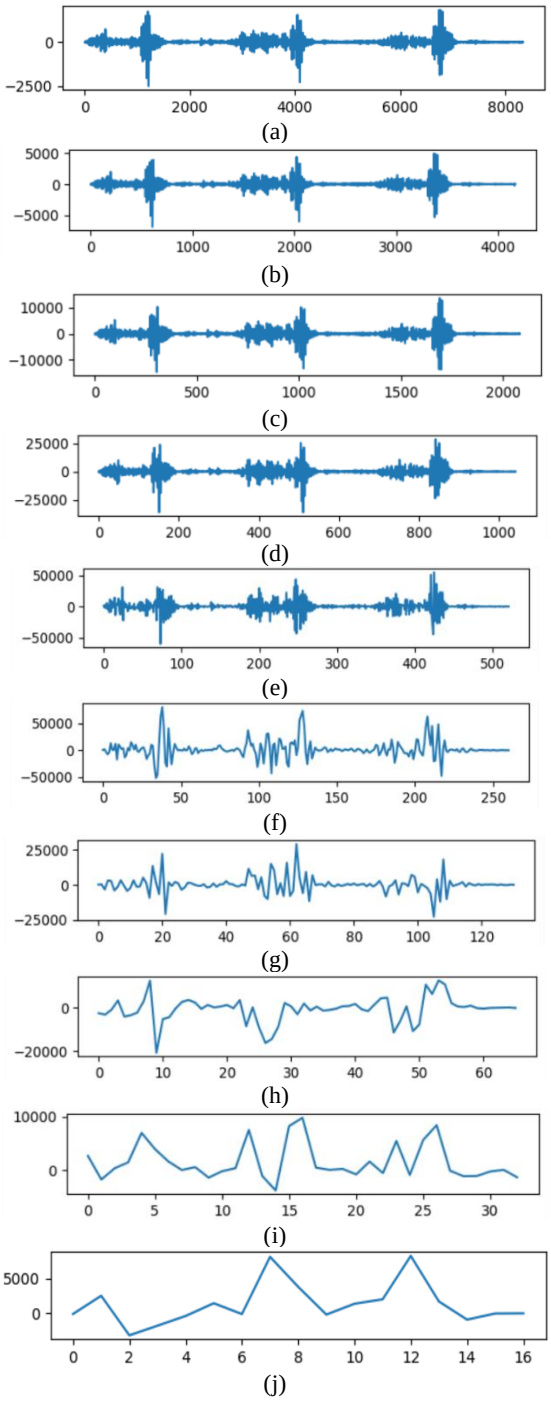


Figure 8. Sample of DWT decomposition Level; (a) Level 1; (b) Level 2; (c) Level 3; (d) Level 4; (e) Level 5; (f) Level 6; (g) Level 7; (h) Level 8; (i) Level 9; (j) Level 10

SFD-CNN Non-Tuning

This section details the training and testing process using a combination of DB1 and DL1. The fine-tuning SFD-CNN model employs the architecture outlined in Table 2, with the parameters specified in Table 3. Tables 4-8 show the output evaluations of the SFD-CNN non-tuning training results.

The three highest accuracies are DB1&DL1 (98%), DB1&DL2 (98%), and DB1&DL3 (98%). Tests were conducted to compare the evaluations of the three combinations.

Table 4. DWT Decomposition level model SFD-CNN Non-Tuning Accuracy (%)

Level	DB 1	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10
1	98	93	75	54	52	51	44	53	51	61
2	98	94	84	85	65	55	53	46	48	43
3	98	93	96	94	82	86	78	73	72	80
4	86	87	82	82	86	81	75	84	76	89
5	88	86	81	81	84	86	84	81	84	88
6	83	84	83	82	89	84	82	86	83	86
7	81	81	87	84	83	82	85	83	76	79
8	75	73	64	55	56	57	47	42	54	44
9	62	56	41	57	44	47	33	38	22	30
10	41	21	22	25	23	32	38	35	36	29

Table 5 shows the average of each fold. It is evident that the DB1&DL1 combination is very slightly different from DB1&DL2, but the DB1&DL1 combination is still superior to the other combinations.

Table 5. Average each fold results of SFD-CNN Non-Tuning (%)

Model	Metrics				
	Accuracy	Precision	Recall	Specificity	F1-Score
SFD-CNN NT DB1&DL1	97.37	93.82	93.44	98.35	93.43
SFD-CNN NT DB1&DL2	97.37	93.35	93.44	98.35	93.42
SFD-CNN NT DB1&DL3	96.21	90.96	90.52	97.63	90.39

This outcome is noteworthy as it reflects the consistency and high performance of the model during testing on that particular fold. The effectiveness of this combination indicates its potential importance in addressing challenges and variations within the dataset.

Tables 6-8 show the performance rates for each class of SFD-CNN non-tuning with different combinations. Table 6 shows the combination of DB1&DL1, using Fold 2, due to its impressive average accuracy of 98.66%. Table 7 shows that the combination of DB1&DL2 also achieves a remarkable accuracy of 98.66% in fold 2, consistent with the result obtained with DB1&DL1. However, there are differences when examining the class-specific accuracies in Fold 2. Meanwhile, Table 8 shows that the combination of DB1&DL3 achieves an accuracy of 96.08 in fold 2.

Although DB1&DL1 performs well in training accuracy across the k-fold, it is essential to understand the subtle differences between combinations and folds for optimal use in real-world scenarios.

Table 6. Performance rate for each class SFD-CNN Non-Tuning DB1&DL1 (%)

Fold		Metrics				
		Accuracy	Precision	Recall	Specificity	F1-Score
1	AS	96.66	89.47	94.44	97.22	91.89
	MR	95.55	91.17	86.11	97.91	88.57
	MS	93.33	77.27	94.44	93.05	85.00
	MVP	97.22	100	86.11	100	92.53
	N	97.22	96.96	88.88	99.30	92.75
2	AS	99.44	100	100	100	98.63
	MR	97.22	97.29	94.44	99.30	93.15
	MS	98.88	91.89	100	97.91	97.29
	MVP	97.77	94.73	88.88	98.61	94.11
	N	100	100	100	100	100
3	AS	97.77	94.44	100	100	94.44
	MR	96.11	93.93	94.44	98.61	89.85
	MS	97.22	89.74	86.11	97.22	93.33
	MVP	98.88	97.22	97.22	99.30	97.22
	N	100	100	100	100	100
4	AS	97.22	91.89	94.44	100	93.15
	MR	93.88	87.87	80.55	97.91	84.05
	MS	94.44	80.95	94.44	97.22	87.17
	MVP	96.11	96.77	83.33	94.44	89.55
	N	99.44	97.29	100	99.30	98.63
5	AS	99.44	97.29	94.59	99.30	98.63
	MR	97.22	87.80	100	96.52	93.50
	MS	98.88	97.22	97.22	99.30	97.22
	MVP	96.11	100	80.55	100	89.23
	N	98.33	94.59	97.22	98.61	95.89

Table 7. Performance rate for each class SFD-CNN Non-Tuning DB1&DL2 (%)

Fold		Metrics				
		Accuracy	Precision	Recall	Specificity	F1-Score
1	AS	97.77	94.44	94.44	98.61	94.44
	MR	96.66	94.11	88.88	98.61	91.42
	MS	94.44	78.26	100	93.05	87.80
	MVP	96.66	96.87	86.11	99.30	91.17
	N	97.77	100	88.88	100	94.11
2	AS	98.88	9722	9722	99.30	97.22
	MR	99.44	100	9722	100	98.59
	MS	97.77	9000	100	97.22	94.73
	MVP	97.77	9705	9166	99.30	94.28
	N	99.44	100	9722	100	98.59
3	AS	96.66	89.47	97.22	97.22	91.89
	MR	93.88	85.71	96.52	96.52	84.50
	MS	95.55	88.88	97.22	97.22	88.88
	MVP	97.22	94.28	98.61	98.61	92.95
	N	100	100	100	100	100
4	AS	96.11	89.18	91.66	97.22	90.41
	MR	94.44	90.62	80.55	97.91	85.29
	MS	96.66	85.71	100	95.83	92.30
	MVP	95.00	90.90	83.33	97.91	86.95
	N	100	100	100	100	100
5	AS	97.77	92.10	97.22	97.91	94.59
	MR	100	100	100	100	100
	MS	98.88	94.73	100	98.61	97.29
	MVP	96.66	100	83.33	100	90.90
	N	98.88	94.73	100	98.61	97.29

https://doi.org/10.25077/jnte.v13n1.1184.2024

Table 8. Performance rate for each class SFD-CNN Non-Tuning DB1&DL3 (%)

Fold		Metrics				
		Accuracy	Precision	Recall	Specificity	F1-Score
1	AS	97.22	94.28	91.66	88.88	92.95
	MR	96.66	94.11	88.88	98.61	91.42
	MS	93.33	78.57	91.66	93.75	84.61
	MVP	94.44	90.62	80.55	97.91	85.29
	N	95.00	86.48	88.88	96.52	87.67
2	AS	87.67	97.22	97.22	99.30	97.22
	MR	98.88	92.30	100	97.91	96.00
	MS	98.33	94.59	97.22	98.61	95.89
	MVP	97.22	100	86.11	100	92.53
	N	98.33	94.59	97.22	98.61	95.89
3	AS	96.66	91.66	91.66	99.30	91.66
	MR	95.55	88.88	88.88	97.91	88.88
	MS	96.11	89.18	91.66	97.22	90.41
	MVP	96.66	96.87	86.11	99.30	91.17
	N	98.33	92.30	100	97.91	96.00
4	AS	95.55	85.00	94.44	97.91	89.47
	MR	92.77	84.84	77.77	95.83	81.15
	MS	95.00	84.61	91.66	96.52	88.00
	MVP	92.77	89.65	72.22	95.83	80.00
	N	98.33	92.30	100	97.91	96.00
5	AS	97.22	91.89	94.44	97.91	93.15
	MR	96.11	83.72	100	95.13	91.13
	MS	98.33	97.14	94.44	99.30	95.77
	MVP	92.77	100	63.88	100	77.96
	N	95.55	83.33	97.22	95.13	89.74

The loss graph in Figure 9 shows a sharp decline from epoch 0 to 5, followed by a gradual decrease to below 0.25% at epoch 10, indicating model stability with no signs of overfitting.

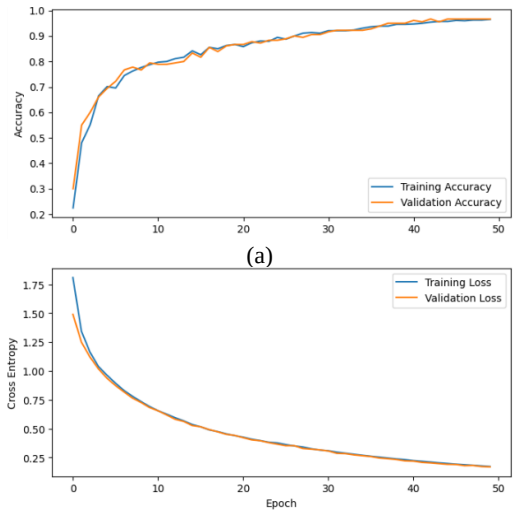


Figure 9. Training and validation SFD-CNN Non-Tuning DB1&DL1 (Fold 2); (a) Accuracy; (b) Loss

Figure 10, which represents the DB1&DL2 combination, closely resembles Figure 9. However, Figure 11, which depicts the DB1&DL3 combination, shows a less stable trend than Figures 9 and 10, although there are no indications of overfitting.

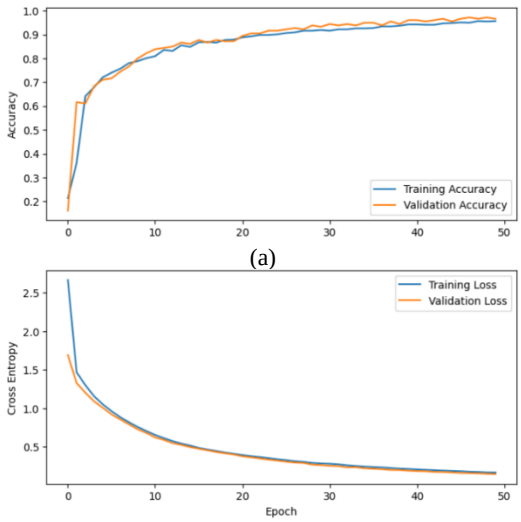


Figure 10. Training and validation SFD-CNN Non-Tuning DB1&DL2 (Fold 2); (a) Accuracy; (b) Loss

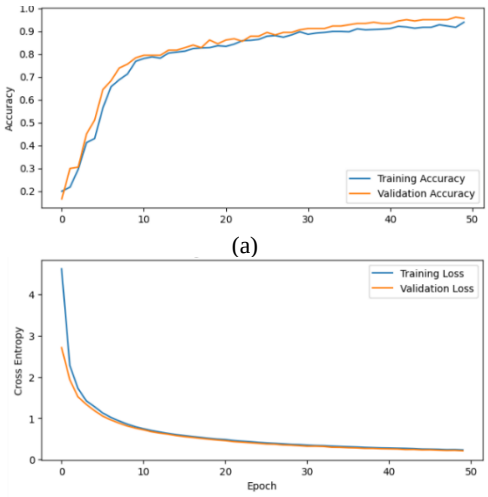


Figure 11. Training and validation SFD-CNN Non-Tuning DB1&DL3 (Fold 2); (a) Accuracy; (b) Loss

Table 9 shows the performance evaluation results of the SFD-CNN non-tuning model on the third-highest decomposition level combination during the testing phase. The highest test result obtained is DB1&DL1, which is 94.80%. Then, the DB1&DL1 combination result is selected for SFD-CNN non-tuning.

Table 9. SFD-CNN Non-Tuning Testing Result (%)

Model	Metrics				
	Accuracy	**Precision**	**Recall**	**Specificity**	**F1-Score**
SFD-CNN NT DB1&DL1	**96.40**	**91.38**	**91.00**	**97.74**	**90.91**
SFD-CNN NT DB1&DL3	96.40	91.66	90.99	97.75	90.86
SFD-CNN NT DB1&DL2	92.80	86.13	82.00	95.50	81.82

Figures 12-14 show the confusion matrix of the SFD-CNN non-tuning model on the third-highest decomposition level combination.

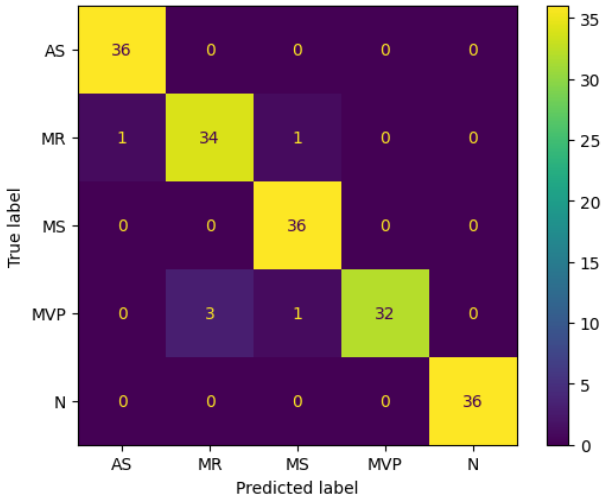


Figure 12. Confusion Matrix SFD-CNN Non-Tuning DB1&DL1 Testing (Fold 2)

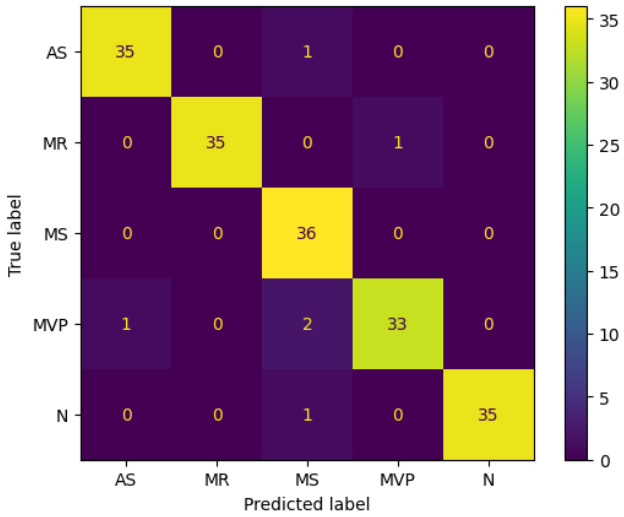


Figure 13. Confusion Matrix SFD-CNN Non-Tuning DB1&DL2 Testing (Fold 2)

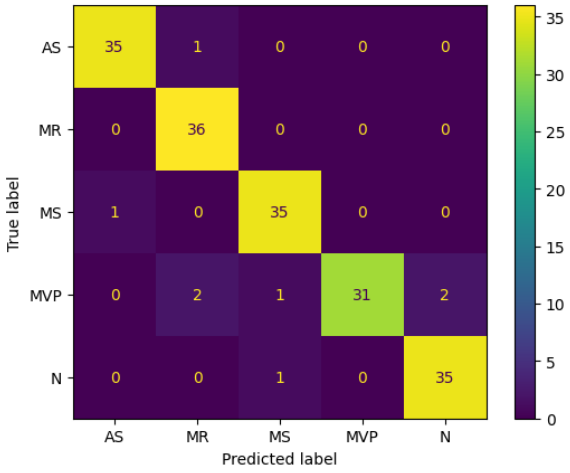


Figure 14. Confusion Matrix SFD-CNN Non-Tuning DB1&DL3 Testing (Fold 2)

SFD-CNN Fine-Tuning

This section details the training and testing process using a combination of DB1 and DL1. The fine-tuning SFD-CNN model employs the architecture outlined in Table 2, with the parameters specified in Table 3. Tables 10-15 below show the output evaluations of the SFD-CNN fine-tuning training results.

Table 10. DWT Decomposition level model SFD-CNN Fine-Tuning Accuracy (%)

Level	DB 1	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10
1	99	93	73	63	65	61	53	51	58	58
2	99	91	91	86	76	63	65	42	55	46
3	99	87	93	92	88	83	81	85	82	79
4	92	88	83	89	83	86	89	90	83	84
5	90	85	94	82	90	83	81	82	87	89
6	88	83	91	88	86	87	85	87	83	88
7	85	76	89	9	88	91	89	85	84	89
8	83	72	65	67	59	51	49	52	48	50
9	74	67	45	49	42	49	43	48	41	39
10	58	43	35	32	36	38	32	34	38	31

Table 11 shows the average of each fold. It is evident that the combination of all three has comparable accuracy, but the DB1&DL1 combination proves to be superior in the SFD-CNN fine-tuning model.

Table 11. Average each fold results of SFD-CNN Fine-Tuning (%)

Model	Metrics				
	Accuracy	Precision	Recall	Specificity	F1-Score
SFD-CNN FT DB1&DL1	98.70	96.91	96.77	99.18	96.76
SFD-CNN FT DB1&DL2	98.53	96.54	96.32	99.08	96.30
SFD-CNN FT DB1&DL3	98.44	96.19	96.10	99.02	96.08

Tables 12-14 present the classification rates for individual classes of the SFD-CNN fine-tuning model with the top three accuracy combinations. Specifically, Table 12 shows the results for the DB1&DL1 combination using fold 2, which has the highest accuracy among the folds at 99.44%. Similarly, Tables 13 and 14 present the results of the fine-tuning process using fold 2, highlighting its consistently high accuracy, recorded at 98.98% and 96.92%, respectively.

These findings demonstrate the effectiveness of fold 2 to achieve superior classification accuracy in the SFD-CNN fine-tuning procedure.

https://doi.org/10.25077/jnte.v13n1.1184.2024

Table 12. Performance rate for each class SFD-CNN Fine-Tuning DB1&DL1 (%)

Fold		Metrics				
		Accuracy	Precision	Recall	Specificity	F1-Score
1	AS	97.22	91.89	100	100	93.15
	MR	97.77	94.44	94.44	97.91	94.44
	MS	99.44	97.29	100	98.61	98.63
	MVP	97.77	97.05	91.66	99.30	94.28
	N	100	100	100	100	100
2	AS	98.88	100	100	100	98.63
	MR	100	97.29	97.22	99.30	98.59
	MS	99.44	100	100	100	97.29
	MVP	98.88	94.73	94.44	98.61	97.14
	N	100	100	100	100	100
3	AS	98.33	97.14	92.10	99.30	95.77
	MR	97.77	94.44	100	98.61	94.44
	MS	98.33	97.14	94.44	99.30	95.77
	MVP	97.77	92.10	97.22	97.91	94.59
	N	100	100	100	100	100
4	AS	99.44	98.33	97.22	100	98.59
	MR	97.77	100	88.88	100	94.11
	MS	98.88	94.73	100	98.61	97.29
	MVP	98.33	92.30	100	97.91	96.00
	N	100	100	100	100	100
5	AS	97.22	87.80	100	96.52	93.50
	MR	100	100	100	100	100
	MS	99.44	100	97.22	100	98.59
	MVP	96.66	100	83.33	100	90.90
	N	98.88	94.73	100	98.61	97.29

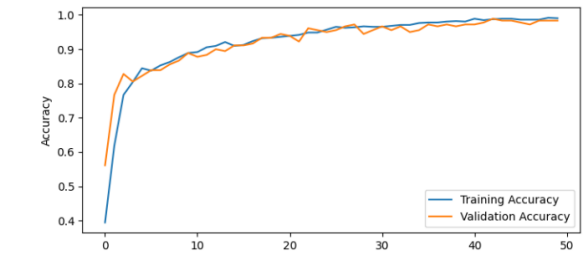
Table 13. Performance rate for each class SFD-CNN Fine-Tuning DB1&DL2 (%)

Fold		Metrics				
		Accuracy	Precision	Recall	Specificity	F1-Score
1	AS	97.77	90.00	100	97.22	94.73
	MR	95.55	88.88	88.88	97.22	88.88
	MS	98.33	97.14	94.44	99.30	95.77
	MVP	97.22	94.28	91.66	98.61	92.95
	N	98.88	100	94.44	100	97.14
2	AS	97.14	97.29	100	99.30	98.63
	MR	99.44	100	94.44	100	97.14
	MS	98.88	94.73	100	98.61	97.29
	MVP	99.44	100	97.22	100	98.59
	N	100	100	100	100	100
3	AS	98.88	97.22	97.22	99.30	97.22
	MR	96.11	100	80.55	100	89.23
	MS	96.66	87.50	97.22	96.52	92.10
	MVP	98.88	94.73	100	98.61	97.29
	N	99.44	97.29	100	99.30	98.63
4	AS	99.44	97.29	100	99.30	98.63
	MR	97.22	100	86.11	100	92.53
	MS	97.22	90.00	100	97.22	94.73
	MVP	99.44	100	972.2	100	98.59
	N	99.44	97.29	100	99.30	98.63
5	AS	97.77	90.00	100	97.22	94.73
	MR	100	100	100	100	100
	MS	100	100	100	100	100
	MVP	97.77	100	88.88	100	94.11
	N	100	100	100	100	100

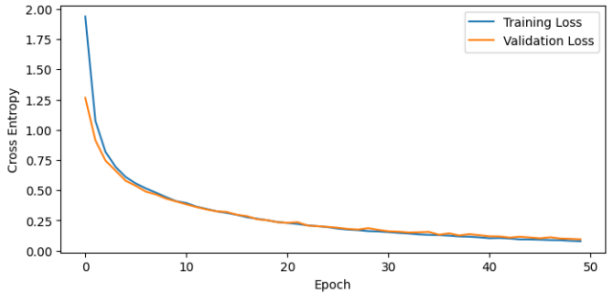
Table 14. Performance rate for each class SFD-CNN Fine-Tuning DB1&DL3 (%)

Fold		Metrics				
		Accuracy	Precision	Recall	Specificity	F1-Score
1	AS	98.33	94.59	97.14	94.44	95.89
	MR	96.11	85.36	97.22	98.61	90.90
	MS	95.55	91.17	86.11	95.83	88.57
	MVP	97.22	96.96	88.88	97.91	92.75
	N	98.33	97.14	94.44	99.30	95.77
2	AS	92.75	94.59	97.22	98.61	95.89
	MR	95.77	97.29	100	99.30	98.63
	MS	98.33	100	97.22	100	98.59
	MVP	99.44	97.14	94.44	99.30	95.77
	N	98.33	100	100	100	100
3	AS	95.77	94.44	94.44	100	94.44
	MR	100	94.59	97.22	98.61	95.89
	MS	97.77	97.14	94.44	99.30	95.77
	MVP	98.33	100	97.22	100	98.59
	N	99.44	97.29	100	99.30	98.63
4	AS	99.44	97.29	100	100	98.63
	MR	96.11	91.42	88.88	99.30	90.14
	MS	99.44	97.29	100	97.91	98.63
	MVP	97.22	96.96	88.88	99.30	92.75
	N	98.88	94.73	100	98.61	97.29
5	AS	98.33	97.77	100	97.91	96.00
	MR	99.44	100	100	99.30	98.63
	MS	100	92.30	100	100	100
	MVP	97.77	97.29	88.88	100	94.11
	N	100	100	100	100	100

Figures 15-17 show the performance of the SFD-CNN fine-tuning model on the training accuracy and validation loss graphs. In all three combinations, there is a significant improvement in the training data at epoch 5, which reaches the 80% to 90% range. The increase continues slowly, and by the end of epoch 40 to 50, shows very little growth. On the loss graph, all three combinations exhibit a decrease in loss at epoch 5. By epoch 50, the loss value has dropped below 0.25%, indicating that all three combinations have stable models.

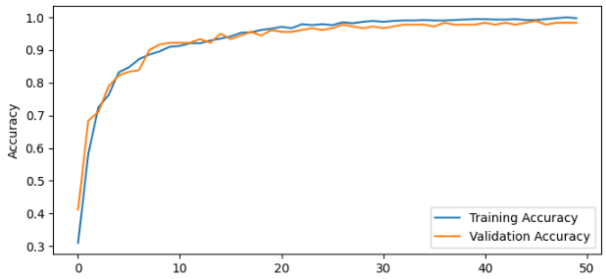


(a)

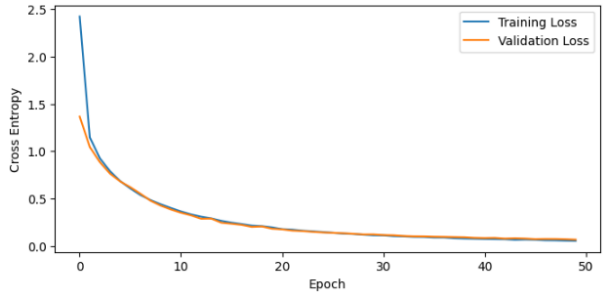


(b)

Figure 15. Training and validation SFD-CNN Fine-Tuning DB1&DL1 (Fold 2); (a) Accuracy; (b) Loss

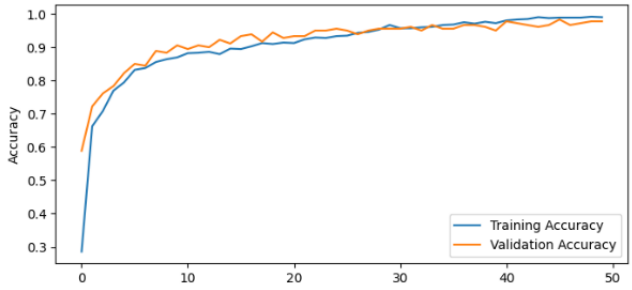


(a)

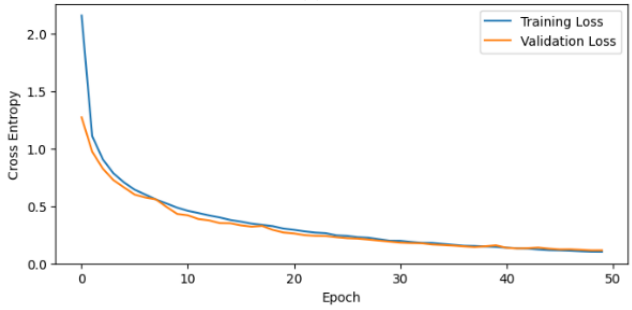


(b)

Figure 16. Training and validation SFD-CNN Fine-Tuning DB1&DL2 (Fold 2); (a) Accuracy; (b) Loss



(a)



(b)

Figure 17. Training and validation SFD-CNN Fine-Tuning DB1&DL3 (Fold 2); (a) Accuracy; (b) Loss

Table 15 shows the results for the third highest decomposition level combination during the test phase in the performance evaluation of the SFD-CNN fine-tuning model.

The DB1&DL1 combination exhibits the highest performance, with a test result of 96.80%. Additionally, the DB1&DL2 combination performs similarly to the DN1&DL1 combination, with only a 0.80% difference.

The combination of DB1&DL3 achieved the lowest result at 95.20%, which was only 1% lower than the other combinations. As a result of its proven proficiency and robustness, this combination was selected for further refinement in the SFD-CNN fine-tuning process.

Table 15. SFD-CNN Fine-Tuning Testing Results (%)

Model	Metrics				
	Accuracy	Precision	Recall	Specificity	F1-Score
SFD-CNN FT DB1&DL1	**96.80**	**93.25**	**91.99**	**98.00**	**92.09**
SFD-CNN FT DB1&DL2	96.00	90.46	90.00	97.50	89.86
SFD-CNN FT DB1&DL3	95.20	90.33	88.00	97.00	88.31

Figures 18-20 show the confusion matrix of the SFD-CNN fine-tuning model on the third-highest decomposition level combination.

The best prediction is shown in Figure 18, namely the DB1&DL1 combination, while the DB1&DL2 combination has a prediction similar to DB1&DL1 but with a few incorrect predictions. The lowest prediction is in the DB1&DL3 combination, which is inferior to the other combinations.

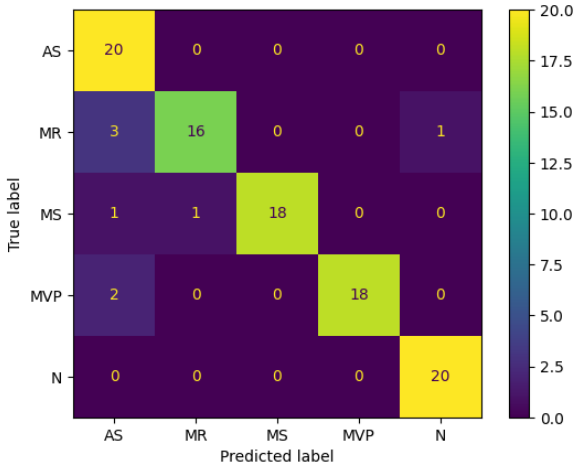


Figure 18. Confusion Matrix SFD-CNN Fine-Tuning DB1&DL1 Testing (Fold 2)

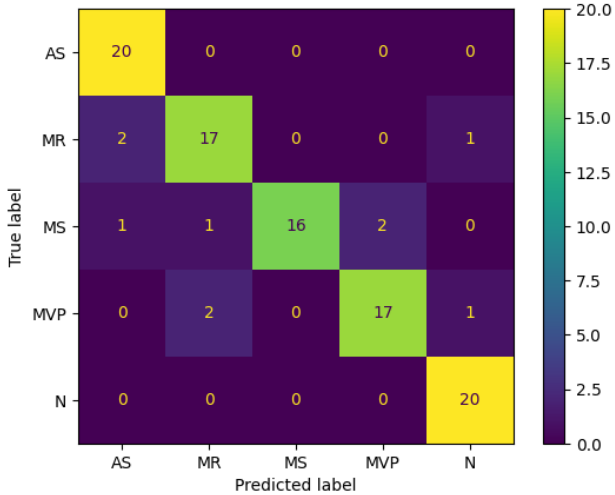


Figure 19. Confusion Matrix SFD-CNN Fine-Tuning DB1&DL2 Testing (Fold 2)

https://doi.org/10.25077/jnte.v13n1.1184.2024

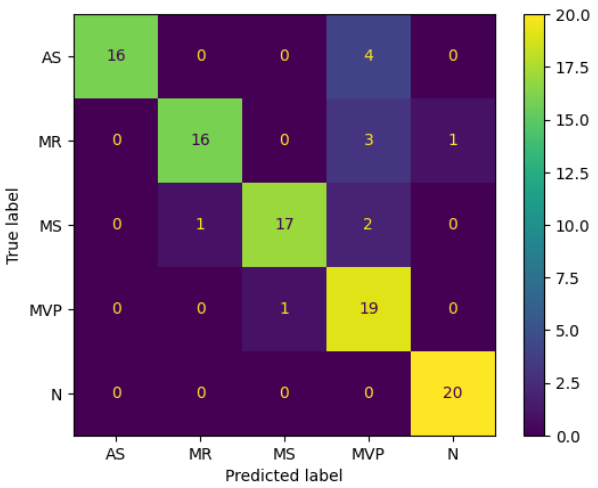


Figure 20. Confusion Matrix SFD-CNN Fine-Tuning DB1&DL3 Testing (Fold 2)

Comparison with SFD-CNN Non-Tuning and Fine-Tuning Model

Table 16 shows the comparison results between the SFD-CNN Non-Tuning and Fine-Tuning tests. In this table, it can be seen that the SFD-CNN Fine-Tuning achieves high scores and outperforms the SFD-CNN Non-Tuning. Therefore, based on the evaluated criteria, the SFD-CNN Fine-Tuning model can be used as a reference for the best model.

Table 16. Comparison SFD-CNN Non-Tuning and Fine-Tuning testing result (%)

Model	Metrics				
	Accuracy	Precision	Recall	Specificity	F1-Score
SFD-CNN FT	**96.80**	**93.25**	**91.99**	**98.00**	**92.09**
SFD-CNN NT	96.40	91.38	91.00	97.74	90.91

Discussion

This subsection discusses the results obtained with our proposed model. In this research, the SFD-CNN fine-tuning model utilized fold two because it has the highest average accuracy compared to other folds, achieving 96.80%. Similarly, the SFD-CNN non-tuning model utilized Fold 2 with the highest average of 96.40%

This study has limitations, including comparing our proposed model with classical machine learning models and the limitations of the dataset used. Furthermore, there is no direct application implementation or a simple prototype that can provide a better understanding of the practical applicability of the proposed model in a real-world environment.

Table 6 shows the results of 100% accuracy on specific classes, where Nti et al. [33] explained that the use of k-fold with small data can increase the complexity of training, which causes the relevance of the features used to predict the class and the model can quickly learn the relationship between features and class labels.

Tables 4 and 10 show the accuracy of SFD-CNN fine-tuning and non-tuning prediction for all Daubechies (DB) combinations and decomposition levels (DL). For the selected features DB1 and DL1, we selected DB1 because it has the highest average accuracy and the selected DL1. After all, it has the highest accuracy. Accuracy combined DL1 tends to be higher because the signal data is still like the original; Chowdhury et al. [34] explain that at each level of decomposition, the signal will be divided into low and high-frequency bands, and the time resolution will be half. This is what makes DL2 and DL3 signals have low accuracy results.

To prove that the increasing level of decomposition affects the accuracy degradation, in Table 12, we compare the three best combinations of DB and DL from each non-tuning and fine-tuning model. However, DL2 and DL3 have lower accuracy than DL1.

In Table 10, the result of SFD-CNN fine-tuning DB1&DL1 and DB1&DL2 is the same, but the accuracy result in Table 16 is different because in DL2, the signal has been decomposed, but not so much; therefore, the accuracy difference is only 0.20%.

The best-selected feature results are DB1&DL1 SFD-CNN non-tuning and DB1&DL1 SFD-CNN fine-tuning. Table 16 compares the testing results of the two proposed models. SDF-CNN fine-tuning achieved the highest accuracy at 96.80%. SDF-CNN fine-tuning uses different parameters than non-tuning, including differentiation of batch size. According to Yu et al. [27], tuning the batch size determines the convergence speed on accuracy improvement.

Comparison with State-of-the-art Related Work

Table 17. Compare existing model approaches

Author	Model	Method	Accuracy	F1-Score
Khade et al. [8]	KNN	MFCC, ZCR, Peak amplitude	90.47	77.33
Khade et al. [8]	AdaBoost	MFCC, ZCR, Peak amplitude	92.85	84.55
Ghosh et al. [12]	RF	Time-frequency features	93.91	-
Khade et al. [8]	SVM	MFCC, ZCR, Peak amplitude	94.07	85.16
Tuncer et al. [13]	DT	PCG-TEP, INCA	95.10	95.14
Proposed Research	**SFD-CNN Fine-Tuning**	**Selected DWT, MFCC, CQT, STFT**	**96.80**	**92.09**

Table 17 in this subsection compares our proposed SFD-CNN model with the state-of-the-art models proposed by other researchers. Our research shows that our model evaluation is superior to theirs. In this comparison, the model presented in Table 17 uses the same dataset as our proposed model. Our model

outperforms other researchers' classic machine learning models and EL algorithms.

According to Taye M. [35], CNN models are considered superior due to their greater complexity compared to classic machine learning. Additionally, CNN models are superior due to their greater complexity than classic machine learning. Additionally, CNN models can self-correct, while classical machine learning still requires human intervention to correct its mistakes. Therefore, our proposed model contributes significantly not only to this research but also to related industries. The results of this study offer valuable insights to the research community and open up opportunities for further development in this area.

CONCLUSIONS

This study proposes an SFD-CNN model for identifying Valve Heart Disease (VHD) types using PCG signals, including AS, MR, MVP, MR, and N cases. The SFD-CNN model fine-tuning at fold 2 showed the highest accuracy for VHD detection, resulting in an accuracy of 96.80%, precision of 93.25%, sensitivity of 91.99%, specificity of 98.00%, and F1-score of 92.09%. Our model outperformed the classic machine learning model and EL algorithm. As a suggestion for future research, consider using alternative datasets or incorporating data augmentation from similar sources. This can enhance the diversity of training data, optimize model performance, and expand its application to patients in a medical setting.

REFERENCES

[1] G. Santangelo *et al.*, "The Global Burden of Valvular Heart Disease: From Clinical Epidemiology to Management," *Journal of Clinical Medicine*, vol. 12, no. 6. Multidisciplinary Digital Publishing Institute (MDPI), March 01, 2023. doi: 10.3390/jcm12062178.

[2] M. Tung, G. Nah, J. Tang, G. Marcus, and F. N. Delling, "Valvular disease burden in the modern era of percutaneous and surgical interventions: The UK Biobank," *Open Heart*, vol. 9, no. 2, Sep. 2022, doi: 10.1136/openhrt-2022-002039.

[3] Y. Coulibaly, A. A. I. Al-Kilany, M. S. A. Latiff, G. Rouskas, S. Mandala, and M. A. Razzaque, "Secure burst control packet scheme for Optical Burst Switching networks," in *2015 IEEE International Broadband and Photonics Conference (IBP)*, IEEE, 2015, pp. 86–91. doi: 10.1109/IBP.2015.7230771.

[4] G. D. Clifford *et al.*, "Classification of Normal/Abnormal Heart Sound Recordings: the PhysioNet/Computing in Cardiology Challenge 2016."

[5] M. Alkhodari and L. Fraiwan, "Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings," *Comput Methods Programs Biomed*, vol. 200, p. 105940, Mar. 2021, doi: 10.1016/J.CMPB.2021.105940.

[6] Yaseen, G. Y. Son, and S. Kwon, "Classification of Heart Sound Signal Using Multiple Features," *Applied Sciences 2018, Vol. 8, Page 2344*, vol. 8, no. 12, p. 2344, Nov. 2018, doi: 10.3390/APP8122344.

[7] M. Yaumil, I. #1, S. Mandala, and M. Pramudyo, "Study of Denoising Method to Detect Valvular Heart Disease Using Phonocardiogram (PCG)," *Indonesia Journal on Computing (Indo-JC)*, vol. 7, no. 1, pp. 31–38, Apr. 2022, doi: 10.34818/INDOJC.2022.7.1.610.

[8] P. J. Khade, P. Mane, S. Mahore, and K. Bhole, "Machine Learning Approach for Prediction of Aortic and Mitral Regurgitation based on Phonocardiogram Signal," in *2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/ICCCNT51525.2021.9579971.

[9] M. Farhan, S. Mandala, and M. Pramudyo, "Detecting Heart Valve Disease Using Support Vector Machine Algorithm based on Phonocardiogram Signal," in *2021 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*, 2021, pp. 128–132. doi: 10.1109/ICICyTA53712.2021.9689142.

[10] W. R. Putra, S. Mandala, and M. Pramudyo, "Study of Feature Extraction Methods to Detect Valvular Heart Disease (VHD) Using a Phonocardiogram," in *2021 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*, 2021, pp. 122–127. doi: 10.1109/ICICyTA53712.2021.9689119.

[11] V. Arora, R. Leekha, R. Singh, and I. Chana, "Heart sound classification using machine learning and phonocardiogram," *https://doi.org/10.1142/S0217984919503214*, vol. 33, no. 26, Sep. 2019, doi: 10.1142/S0217984919503214.

[12] S. K. Ghosh, R. K. Tripathy, R. N. Ponnalagu, and R. B. Pachori, "Automated Detection of Heart Valve Disorders from the PCG Signal Using Time-Frequency Magnitude and Phase Features," *IEEE Sens Lett*, vol. 3, no. 12, Dec. 2019, doi: 10.1109/LSENS.2019.2949170.

[13] T. Tuncer, S. Dogan, R. S. Tan, and U. R. Acharya, "Application of Petersen graph pattern technique for automated detection of heart valve diseases with PCG signals," *Inf Sci (N Y)*, vol. 565, pp. 91–104, Jul. 2021, doi: 10.1016/j.ins.2021.01.088.

[14] S. Mandala, Y. N. Fuadah, M. Arzaki, and F. E. Pambudi, *Performance Analysis of Wavelet-Based Denoising Techniques for ECG Signal*. 2017.

[15] R. I. Aljohani, H. A. Hosni Mahmoud, A. Hafez, and M. Bayoumi, "A Novel Deep Learning CNN for Heart Valve Disease Classification Using Valve Sound Detection," *Electronics 2023, Vol. 12, Page 846*, vol. 12, no. 4, p. 846, Feb. 2023, doi: 10.3390/ELECTRONICS12040846.

[16] A. M. Alqudah, S. Qazan, and A. Alqudah, "Detection of Valvular Heart Diseases Using Fourier Transform and Simple CNN Model," *Article in IAENG International Journal of Computer Science*, Accessed: April 28, 2023. [Online]. Available: https://www.researchgate.net/publication/365993591

[17] J. Li, L. Ke, and Q. Du, "Classification of heart sounds based on thewavelet fractal and twin support vector machine," *Entropy*, vol. 21, no. 5, May 2019, doi: 10.3390/e21050472.

[18] N. Binta, I. Sabur, K. Nuhash, and T. Hasan, "Hilbert-Envelope Features for Cardiac Disease Classification

https://doi.org/10.25077/jnte.v13n1.1184.2024

from Noisy Phonocardiograms", doi: 10.1101/2020.11.17.20233064.

[19] J. Karhade, S. Dash, S. K. Ghosh, D. K. Dash, and R. K. Tripathy, "Time-Frequency-Domain Deep Learning Framework for the Automated Detection of Heart Valve Disorders Using PCG Signals," *IEEE Trans Instrum Meas*, vol. 71, 2022, doi: 10.1109/TIM.2022.3163156.

[20] T. Sinha Roy, J. K. Roy, and N. Mandal, "Conv-Random Forest-Based IoT: A Deep Learning Model Based on CNN and Random Forest for Classification and Analysis of Valvular Heart Diseases," *IEEE Open Journal of Instrumentation and Measurement*, vol. 2, pp. 1–17, Sep. 2023, doi: 10.1109/ojim.2023.3320765.

[21] T. S. Roy, J. K. Roy, and N. Mandal, "Classifier identification using deep learning and machine learning algorithms for the detection of valvular heart diseases," *Biomedical Engineering Advances*, vol. 3, p. 100035, Jun. 2022, doi: 10.1016/J.BEA.2022.100035.

[22] S. I. Flores-Alonso, B. Tovar-Corona, and R. Luna-García, "Deep Learning Algorithm for Heart Valve Diseases Assisted Diagnosis," *Applied Sciences 2022, Vol. 12, Page 3780*, vol. 12, no. 8, p. 3780, Apr. 2022, doi: 10.3390/APP12083780.

[23] A. M. Alqudah, H. Alquran, and I. A. Qasmieh, "Classification of heart sound short records using bispectrum analysis approach images and deep learning," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 9, no. 1, Dec. 2020, doi: 10.1007/s13721-020-00272-5.

[24] S. A. Singh, T. G. Meitei, and S. Majumder, "Short PCG classification based on deep learning," in *Deep Learning Techniques for Biomedical and Health Informatics*, Elsevier Inc., 2020, pp. 141–164. doi: 10.1016/B978-0-12-819061-6.00006-9.

[25] Ö. Arslan and M. Karhan, "Effect of Hilbert-Huang transform on classification of PCG signals using machine learning," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pp. 9915–9925, Nov. 2022, doi: 10.1016/J.JKSUCI.2021.12.019.

[26] Amelia F and Gunawan D, "DWT-MFCC Method for Speaker Recognition System with Noise," in *2019 7th International Conference on Smart Computing & Communications (ICSCC).*, 2019. doi: 10.1109/ICSCC.2019.8843660.

[27] S. Mandala, Y. N. Fuadah, M. Arzaki, and F. E. Pambudi, *Performance Analysis of Wavelet-Based Denoising Techniques for ECG Signal*. 2017.

[28] J. Leo, C. Loong, K. S. Subari, M. K. Abdullah, N. Ahmad, and R. Besar, "Comparison of MFCC and Cepstral Coefficients as a Feature Set for PCG Biometric Systems."

[29] P. Singh, S. Waldekar, M. Sahidullah, G. Saha, and S. Goutam, "Analysis of constant-Q filterbank based representations for speech emotion recognition Analysis of constant-Q filterbank based representations for speech emotion recognition Analysis of constant-Q filterbank based representations for speech emotion recognition," *Digit Signal Process*, vol. 130, p. 103712, 2022, doi: 10.1016/j.dsp.2022.103712ï.

[30] J. K. Das, A. Ghosh, A. K. Pal, S. Dutta, and A. Chakrabarty, "Urban Sound Classification Using Convolutional Neural Network and Long Short Term Memory Based on Multiple Features," *4th International Conference on Intelligent Computing in Data Sciences, ICDS 2020*, Oct. 2020, doi: 10.1109/ICDS50568.2020.9268723.

[31] H. Jeon, Y. Jung, S. Lee, and Y. Jung, "Area-efficient short-time fourier transform processor for time–frequency analysis of non-stationary signals," *Applied Sciences (Switzerland)*, vol. 10, no. 20, pp. 1–10, Oct. 2020, doi: 10.3390/app10207208.

[32] S. Fitriani, S. Mandala, and M. A. Murti, "Review of semi-supervised method for Intrusion Detection System," in *2016 Asia Pacific Conference on Multimedia and Broadcasting (APMediaCast)*, 2016, pp. 36–41. doi: 10.1109/APMediaCast.2016.7878168.

[33] I. K. Nti, O. Nyarko-Boateng, and J. Aning, "Performance of Machine Learning Algorithms with Different K Values in K-fold CrossValidation," *International Journal of Information Technology and Computer Science*, vol. 13, no. 6, pp. 61–71, Dec. 2021, doi: 10.5815/ijitcs.2021.06.05.

[34] M. T. H. Chowdhury, K. N. Poudel, and Y. Hu, "Automatic Phonocardiography Analysis Using Discrete Wavelet Transform," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Aug. 2019. doi: 10.1145/3387168.3387172.

[35] M. M. Taye, "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions," *Computers*, vol. 12, no. 5. MDPI, May 01, 2023. doi: 10.3390/computers12050091.

AUTHOR(S) BIOGRAPHY

Muhammad Rafli Ramadhan

Was born in Bekasi, Indonesia. He is currently pursuing his bachelor's degree at the School of Computing, Telkom University. His research interests are artificial intelligence and biomedical engineering, specifically audio and image processing.

Satria Mandala

Received the Ph.D. degree in computer science from Universiti Teknologi Malaysia. He is currently the Director of the Human Centric Engineering (Humic Eng) and a member of the School of Computing (SoC), Telkom University, Indonesia. His research interests include issues related to wireless network security, the Internet of Things (IoT), biomedical engineering, digital signal, and image processing.

Rafi Ullah

Received an MS degree in Computer Systems Engineering from Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Pakistan, in 2006 and a PhD degree in Computer and Information Sciences from Pakistan Institute of Engineering and Applied Sciences (PIEAS), Pakistan, in 2010. From 2011 to 2012, he was a Postdoctoral Fellow at Universiti Teknologi PETRONAS. He is currently an Associate Professor at Universiti

Teknologi PETRONAS, Malaysia. His research interests include image/video processing, digital watermarking, cybersecurity, computer vision, machine learning and brain/EEG signals.

Wael M.S. Yafooz

Wael Yafooz is Associate Professor in the computer Science. He received his PhD in Computer Science in 2014 from UiTM. He was invited as a speaker in many international conferences held in Bangladesh, Thailand, India, China and Russia. His research interest includes, Data Mining, Machine Learning, Deep Learning, Natural Language Processing, Social Network Analytics and Data Management.

Muhammad Qomaruddin

Muhammad Qomaruddin obtained his B.Sc. in Informatics Engineering from the Institut Sains Teknologi "AKPRIND", Yogyakarta, Indonesia, and obtained Master's and Ph.D. degrees in Computer Science from UTM Malaysia. He is a senior lecturer at Universitas Islam Sultan Agung (UNISSULA), Semarang, Indonesia. The area of research interest concentrates on Education Technology, Human-computer interaction, Social Computing, Information Systems, and the social impact of technology.

APPENDICES

1. DWT Decomposition level model SFD-CNN Non-Tuning Precision (%)

Level	DB 1	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10
1	96	84	79	52	50	54	48	50	55	51
2	95	92	89	83	66	58	53	47	48	43
3	95	96	91	93	87	85	80	76	78	80
4	84	91	89	89	86	87	87	87	86	89
5	83	88	89	85	87	87	85	87	87	88
6	83	89	87	87	84	87	88	84	86	86
7	81	82	86	87	81	86	80	83	81	79
8	72	70	66	58	58	79	46	44	58	44
9	61	55	47	44	34	44	38	32	34	30
10	44	26	26	32	33	32	31	27	32	36

2. DWT Decomposition level model SFD-CNN Non-Tuning Sensitivity (%)

Level	DB 1	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10
1	96	84	79	52	50	54	48	50	55	51
2	95	92	89	83	66	58	53	47	48	43
3	95	96	91	93	89	85	80	76	78	80
4	84	91	89	89	87	86	87	88	89	89
5	83	88	89	85	87	87	85	85	86	88
6	81	89	87	87	84	86	87	87	86	86
7	77	82	86	87	81	83	82	80	83	79
8	71	70	66	58	58	43	46	46	44	44
9	66	55	47	44	34	37	35	38	32	30
10	52	26	26	32	33	32	31	27	37	41

3. DWT Decomposition level model SFD-CNN Non-Tuning Specificity (%)

Level	DB 1	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10
1	99	84	79	52	50	54	48	50	55	51
2	98	92	89	83	66	58	53	47	48	43
3	98	96	91	93	83	66	58	53	78	80
4	84	91	84	89	87	84	85	80	87	86
5	79	85	86	89	84	82	87	71	82	84
6	78	81	71	85	78	85	85	74	83	82
7	81	85	86	87	74	83	85	70	78	79
8	72	68	66	87	51	63	52	46	44	44
9	61	51	49	44	47	37	35	38	32	30
10	35	32	23	42	43	22	41	27	20	25

4. DWT Decomposition level model SFD-CNN Non-Tuning F1-Score (%)

Level	DB 1	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10
1	96	82	75	61	65	54	52	63	53	52
2	95	91	82	85	66	58	53	47	48	42
3	94	92	89	93	89	85	80	76	78	72
4	84	81	86	89	87	86	73	79	80	89
5	83	85	81	85	87	87	85	85	86	88
6	83	86	85	84	84	86	87	87	86	86
7	81	81	79	72	81	83	82	80	83	71
8	72	69	62	53	58	43	46	46	44	44
9	61	56	39	45	34	37	35	38	32	30
10	44	38	21	29	43	22	21	27	30	29

5. DWT Decomposition level model SFD-CNN Fine-Tuning Precision(%)

Level	DB 1	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10
1	98	94	78	61	68	52	66	49	53	54
2	98	90	91	86	77	62	68	49	60	48
3	97	83	94	94	91	89	83	84	82	80
4	92	87	89	90	88	93	91	92	88	90
5	89	89	90	89	93	88	92	89	91	91
6	93	84	92	92	89	94	92	93	91	90
7	85	72	89	91	89	87	83	85	87	81
8	82	74	73	71	67	62	56	53	48	52
9	71	62	66	64	52	43	31	42	35	31
10	55	39	33	36	37	31	32	28	26	21

6. DWT Decomposition level model SFD-CNN Fine-Tuning Sensitivity (%)

Level	DB 1	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10
1	98	92	73	66	62	54	56	44	52	52
2	98	89	89	89	75	61	68	42	59	42
3	97	81	88	97	89	90	83	86	82	76
4	94	83	82	92	82	91	91	87	81	88
5	91	84	86	91	91	83	92	82	88	90
6	90	85	91	95	86	91	92	87	87	89
7	89	71	86	92	83	83	89	82	82	86
8	81	72	74	78	61	66	49	51	45	52
9	65	59	67	63	49	49	33	39	32	44
10	45	34	32	37	32	38	29	30	27	32

7. DWT Decomposition level model SFD-CNN Fine-Tuning Specificity (%)

Level	DB 1	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10
1	99	91	75	67	61	55	55	43	51	53
2	98	87	90	88	74	62	67	41	57	44

https://doi.org/10.25077/jnte.v13n1.1184.2024

3	97	78	87	96	90	91	82	87	81	75
4	93	82	81	91	83	93	90	85	83	86
5	90	83	85	90	93	82	91	83	86	89
6	89	83	92	94	87	90	94	88	86	91
7	86	71	85	95	84	83	88	83	83	84
8	82	73	73	80	62	65	47	50	47	55
9	67	58	66	65	50	45	35	38	31	43
10	43	35	31	34	33	36	30	32	28	31

8. DWT Decomposition level model SFD-CNN Fine-Tuning F1-score(%)

Level	DB 1	DB 2	DB 3	DB 4	DB 5	DB 6	DB 7	DB 8	DB 9	DB 10
1	98	90	74	68	60	56	56	42	52	54
2	98	88	91	87	74	61	68	43	58	43
3	97	79	88	95	91	90	81	86	82	76
4	92	83	83	92	82	92	89	82	84	85
5	91	84	82	91	94	81	90	84	86	88
6	88	81	90	92	88	91	92	89	87	92
7	87	73	84	94	83	82	86	82	85	83
8	83	77	72	81	63	66	48	51	46	56
9	68	54	65	66	51	47	37	37	32	42
10	44	34	32	35	34	33	33	31	29	32