IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 3, June 2026, pp. 2643~2654
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i3.pp2643-2654

ValveHealthNet: a light deep learning model for accurate valvular heart disorder detection

Ausilah Alfraihat[1], Wafaa Al-Sharu[2], Ali Mohammad Alqudah[3]
[1]Department of Biomedical Engineering, Faculty of Engineering, The Hashemite University, Zarqa, Jordan
[2]Department of Electrical Engineering, Faculty of Engineering, The Hashemite University, Zarqa, Jordan
[3]Independent Researcher, Winnipeg, Canada

Article Info

Article history:

Received Dec 3, 2024
Revised Feb 11, 2026
Accepted Apr 21, 2026

Keywords:

Artificial intelligence
Convolutional neural networks
Deep learning
Embedded healthcare systems
Heart sound analysis
Heart valve
Heart valve diseases

ABSTRACT

Valvular heart disease (VHD) is a significant global health issue, contributing to increased morbidity and mortality rates, particularly in aging populations. Current diagnostic methods, such as echocardiography and manual auscultation, face limitations in accessibility and accuracy, particularly in resource-constrained environments. This study introduces ValveHealthNet, a lightweight deep learning model designed to classify various VHDs using heart sound recordings. Leveraging a dataset of over 10,000 heart sounds, minimal preprocessing was applied by converting the audio signals into power spectra before feeding them into a convolutional neural network (CNN) combined with a bidirectional long short-term memory (BiLSTM) network. This model achieved impressive results, with an accuracy of 98% in training and testing and 98.4% through 10-fold cross-validation. This highly efficient model can be used in embedded systems, providing a cost-effective, AI-driven solution for early detection of VHD in settings where advanced diagnostic tools may be unavailable.

Corresponding Author:

Ali Mohammad Alqudah
Independent Researcher
Winnipeg, MB R2M 3R1, Canada
Email: ali_qudah@hotmail.com

1. INTRODUCTION

Valvular heart disease (VHD) represents a significant clinical burden, especially in aging populations where the incidence continues to rise. Currently, approximately 13.3% of older individuals are affected by VHD, and projections indicate that death rates attributable to VHD may double in the coming 25 years [1], [2]. Frequently, VHD remains undiagnosed until severe complications, such as heart failure, leading to worsened patient outcomes due to the disease's typically asymptomatic progression [3], [4]. As VHD severity increases, healthcare costs rise accordingly, underscoring the need for early detection and timely intervention to both improve patient prognosis and alleviate financial strain on healthcare systems [5]. Despite these pressing concerns, delays in diagnosis are common, often attributed to inadequate clinician use of or proficiency in auscultation [6], [7].

Even among trained medical professionals, the sensitivity and specificity of identifying heart murmurs remain suboptimal, ranging from 35% to 69% [6], [7]. The overall incidence of VHD has surged by 45% in the past 30 years, with an estimated 401 new cases per 100,000 people annually [8]. Although echocardiography has become the gold standard for non-invasive diagnosis of VHD, more than seven million echocardiograms are performed annually in North America, yet many cases still go undetected [9]–[11]. Expanding echocardiographic screening to encompass all at-risk populations is impractical given current

clinical resources and financial constraints [12]. Furthermore, handheld ultrasound devices, while promising, still depend heavily on the operator's skill, thereby introducing variability in diagnostic accuracy [13].

In recent years, artificial intelligence (AI)-enhanced auscultation has emerged as a viable approach to addressing gaps in clinician expertise [14], [15]. Previous research has demonstrated that AI algorithms can distinguish normal from abnormal heart sounds with sensitivities between 78.5% and 88.5% and specificities ranging from 81.5% to 98.33% [8]–[10]. However, existing AI-driven systems have been largely limited to differentiating normal from abnormal sounds without providing a diagnosis of specific valvular conditions. Moreover, studies primarily rely on processed heart sound signals or phonocardiogram (PCG) data rather than raw heart sounds obtained in clinical settings, which diminishes their applicability in real-world clinical practice [16].

Barua *et al.* [17] proposed a handcrafted learning model for VHD detection that integrates a multilevel feature extraction method based on the dual symmetric tree pattern (DSTP) and discrete wavelet transform (DWT). The DSTP operates as a textural feature extractor, similar to the local binary pattern (LBP), by capturing low-level textural features. To extend its capability, the model applies DWT to generate subbands of heart sound signals, and DSTP is repeatedly applied across these subbands to create a multileveled feature vector. This approach allowed for capturing more complex representations of heart sound data, but it also highlights a limitation, as handcrafted features may not be as effective or adaptable as learned features in deep learning models.

Jiang *et al.* [18] recently addressed some of the limitations by developing a deep learning model trained on raw heart sound data. Their model achieved promising sensitivity and specificity rates for detecting various VHD phenotypes, ranging from 71.4% to 100% and 83.5% to 100%, respectively [18]. This advancement marks an important step forward, as the model operates directly on raw clinical heart sound data, removing the need for laborious preprocessing. However, despite these improvements, Jiang's model exhibited reduced sensitivity in detecting mixed valvular diseases and aortic valve lesions [18]. Furthermore, the model's generalizability remains a challenge, as it was primarily validated within inpatient settings, necessitating broader validation across diverse clinical populations. The detection of aortic valve abnormalities continues to pose difficulties, with lower sensitivity observed in these cases, indicating the need for further refinement of deep learning approaches to fully address the complexity of VHD phenotypes.

This study built upon this foundation by developing a lightweight, highly accurate deep learning model for sound classification that leverages a large dataset of 10,366 heart sounds, collected using electronic stethoscopes. By minimizing preprocessing and converting heart sounds into power spectra, a light convolutional neural network (CNN) was applied to extract and select multilevel features. This approach yielded a highly efficient model, achieving 98% accuracy in both training and testing phases, with a 98.4% accuracy upon 10-fold cross-validation.

2. METHOD

All experiments presented in this work were designed and executed to develop, train, and evaluate the proposed CNN–bidirectional long short-term memory (BiLSTM) model for VHD classification using power spectral representations of heart sound signals. The entire experimental pipeline, including data preprocessing, model implementation, training, and validation, was carried out using MATLAB R2024b (MathWorks, Natick, MA, and USA). Computational acceleration was achieved using an NVIDIA GeForce GTX 1650 graphics processing unit (GPU) with 4 GB of dedicated memory, and all experiments were performed on a system equipped with 32 GB of RAM. This setup ensured efficient model training while reflecting a practical and reproducible research environment suitable for deployment-oriented studies.

2.1. Dataset

The heart sound dataset [17] was prospectively collected from patients attending the Cardiology Department at Firat University Hospital, with approval from the non-interventional research ethics board. It includes 10,366 heart sound recordings from 651 subjects, divided into nine VHD classes and one healthy class (Table 1). Each sound was captured at a fixed duration of 2 seconds using the Littmann 3,200 digital stethoscope at a sampling frequency of 8 kHz. The diagnostic class for each subject was confirmed through prior echocardiography. This dataset is publicly available and can be accessed [17].

2.2. Data preprocessing and power spectrum analysis

The PCG signals were transformed into the frequency domain using a welch-based power spectral density (PSD) estimation, which is appropriate given the quasi-stationary characteristics of heart sound signals. As implemented in the code, the signals were sampled at 8 kHz and segmented using a Hamming window of 1,024 samples with 50% overlap. For each segment, the periodogram was computed via the fast

Fourier transform (FFT), and the resulting power spectra were averaged across segments to obtain a robust PSD estimate for each recording.

Following PSD estimation, only the first 200 frequency bins were retained, corresponding to the low-frequency band (0–200 Hz) that contains the most diagnostically relevant heart sound components, including S1, S2, and pathological murmurs. This frequency range has been shown to capture clinically meaningful spectral differences in PCG signals with high confidence [19], [20]. Each resulting PSD vector was then formatted as a sequence input of 200 and used directly for training, validation, and testing, reflecting a minimal and computationally efficient preprocessing pipeline.

Regarding class imbalance, the dataset used in this study exhibited a relatively balanced distribution across diagnostic classes, and therefore no explicit imbalance correction techniques (such as class-weighted loss functions or data resampling) were applied during training. Nevertheless, the model was evaluated using separate training, validation, and test sets to ensure unbiased performance estimation. While this work focuses on balanced data to establish baseline performance, future studies will investigate the robustness of the proposed model under clinically realistic imbalanced conditions, including the use of weighted loss functions and oversampling strategies to mitigate class dominance and enhance generalizability.

Table 1. Distribution of subjects and heart sounds by diagnostic class

Class	Number of subjects	Number of sounds	Percentage of total heart sounds (%)
Severe aortic stenosis	23	999	9.6
Mild aortic stenosis	54	738	7.1
Mild mitral stenosis	45	656	6.3
Moderate mitral stenosis	69	1035	10.0
Severe mitral regurgitation	90	1338	12.9
Mild mitral regurgitation	72	1077	10.4
Moderate mitral regurgitation	54	792	7.6
Severe tricuspid	72	1077	10.4
Moderate tricuspid	72	1080	10.4
Healthy subjects	108	1614	15.6

2.3. Model architecture and hyperparameter settings

The proposed network follows a CNN–BiLSTM architecture implemented using MATLAB's layerGraph framework. The input to the model is a sequence input layer of size 200/times1, corresponding to the power spectral representation of each heart sound segment. A sequence folding layer is employed to enable convolutional processing of sequential data. The convolutional backbone consists of four two-dimensional convolutional layers with kernel sizes of 3/times1 and filter depths of 48, 32, 16, and 32, respectively. Each convolutional layer is followed by batch normalization and rectified linear unit (ReLU) activation to stabilize training and introduce nonlinearity. Three max-pooling layers with a pool size of 2 and a stride of 2 are used to progressively reduce the temporal resolution and extract hierarchical features. After convolutional feature extraction, a sequence unfolding layer restores the temporal structure, followed by a flatten layer [21]. Figure 1 illustrates the proposed ValveHealthNet architecture, which integrates CNNs with a BiLSTM layer for heart sound classification. Power spectral features are first processed through multiple convolutional blocks consisting of convolution, batch normalization, ReLU activation, and max-pooling layers to extract hierarchical spectral representations. The sequence folding and unfolding layers enable CNN-based processing of sequential data, while the BiLSTM layer captures temporal dependencies across the extracted features. The final fully connected and softmax layers perform multi-class classification across ten diagnostic categories.

Temporal dependencies are modeled using a BiLSTM layer with 64 hidden units and an output mode set to last, allowing the network to capture both past and future contextual information in the sequence. The final classification stage consists of a fully connected layer with 10 output neurons, corresponding to the target classes, followed by a softmax layer and a classification layer for final prediction. The detailed layer configuration and corresponding hyperparameters are summarized in Table 2.

2.4. Training and validation

The model was trained and validated using the processed dataset. Training was performed using the Adam optimizer with an initial learning rate of 0.001 for a maximum of 100 epochs and a mini-batch size of 10, with the training data shuffled at each epoch to reduce bias and improve generalization. Model training was executed on a GPU to accelerate convergence. No explicit early stopping criterion was applied; instead, training was carried out for the full number of epochs while monitoring convergence through the training-progress plots provided by MATLAB. These hyperparameter choices were selected empirically to balance classification performance, training stability, and computational efficiency, particularly with the goal

of supporting deployment on resource-constrained and embedded platforms. After training, the model's performance was evaluated using k-fold cross-validation on a separate test set. Figure 2 illustrates the k-fold cross-validation process adopted in this study.

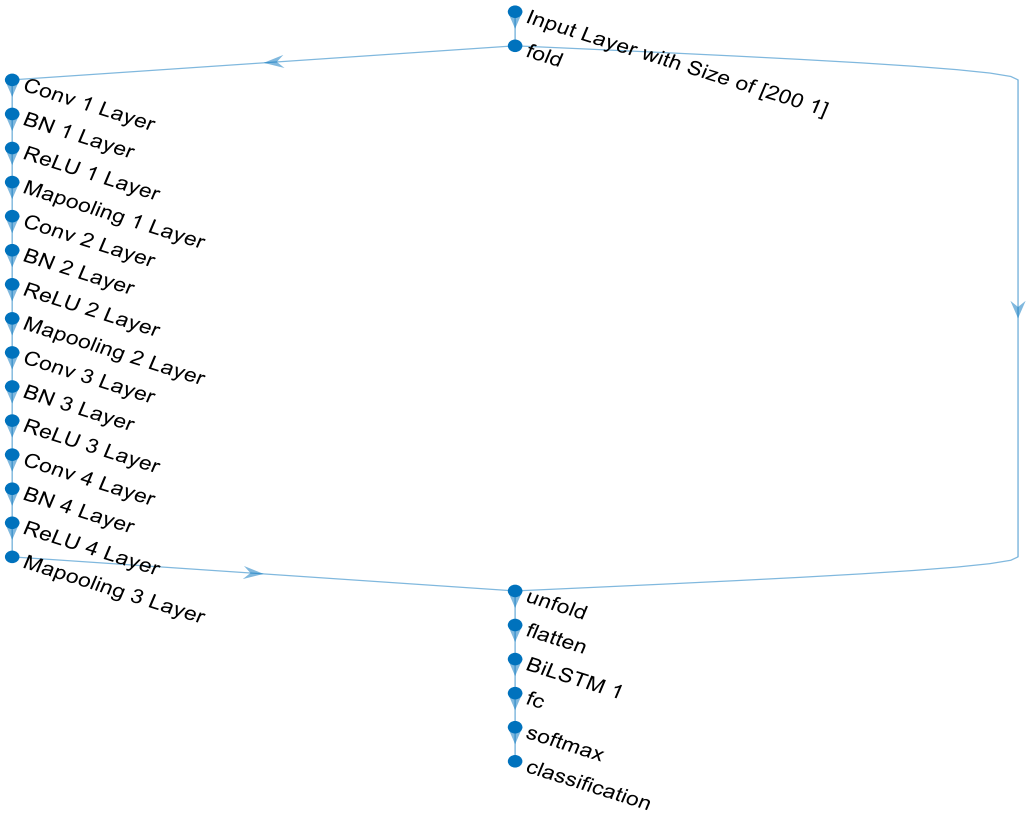

Figure 1. CNN–BiLSTM model architecture for VHD classification

Table 2. Detailed architecture and hyperparameter settings of the proposed CNN–BiLSTM model

Layer type	Layer name	Parameters
Sequence input	Input layer	Input size: (200\times 1\times 1)
Sequence folding	Folding layer	Enables CNN processing of sequence data
Convolution	Conv 1	48 filters, kernel: (3 \times 1), padding: same
Batch normalization	BN 1	—
Activation	ReLU 1	ReLU
Max pooling	MaxPool 1	Pool size: 2, stride: 2
Convolution	Conv 2	32 filters, kernel: (3 \times 1), padding: same
Batch normalization	BN 2	—
Activation	ReLU 2	ReLU
Max pooling	MaxPool 2	Pool size: 2, stride: 2
Convolution	Conv 3	16 filters, kernel: (3 \times 1), padding: same
Batch normalization	BN 3	—
Activation	ReLU 3	ReLU
Convolution	Conv 4	32 filters, kernel: (3 \times 1), padding: same
Batch normalization	BN 4	—
Activation	ReLU 4	ReLU
Max pooling	MaxPool 3	Pool size: 2, stride: 2
Sequence unfolding	Unfolding layer	Restores temporal structure
Flatten	Flatten layer	Feature vector conversion
Bilstm	BiLSTM	64 hidden units, output mode: last
Fully connected	FC	10 neurons
Softmax	Softmax	Class probability estimation

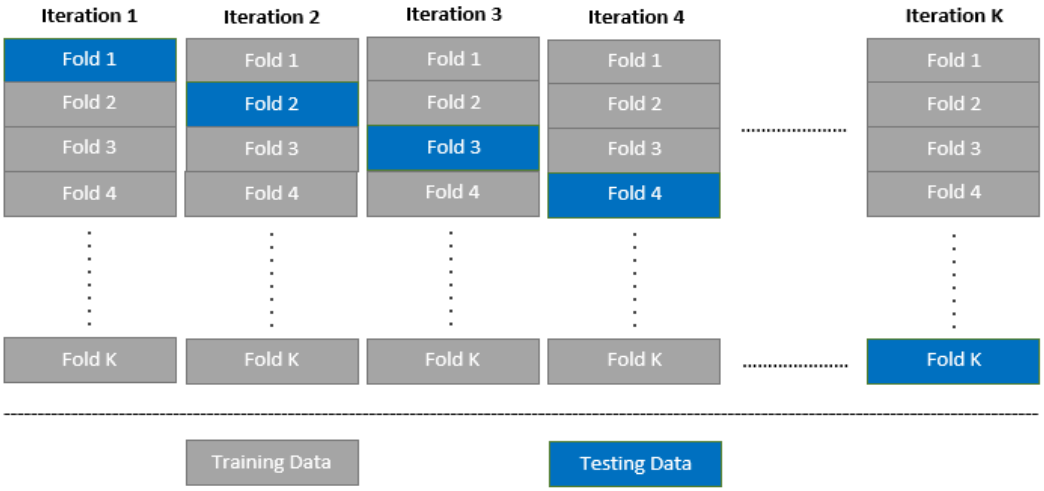

Figure 2. K-fold cross-validation process applied in our methodology

3. RESULTS AND DISCUSSION

3.1. Data preprocessing and power spectrum analysis

Figure 3 illustrates the power spectrum of heart sounds for each VHD category, with 95% confidence intervals. This figure presents the average PSD of heart sound recordings for each VHD category and healthy subjects, along with 95% confidence intervals. The analysis focuses on the 0–200 Hz frequency range, which contains diagnostically relevant components such as S1, S2, and pathological murmurs. Distinct spectral patterns can be observed between different valve pathologies, such as aortic stenosis and mitral regurgitation, providing physiological justification for the feature representation used by the model. These spectral differences form the basis for CNN's discriminative feature learning. The analysis highlights critical differences in the frequency domain between conditions like severe aortic stenosis and mitral regurgitation, demonstrating the basis for the model's feature extraction and classification.

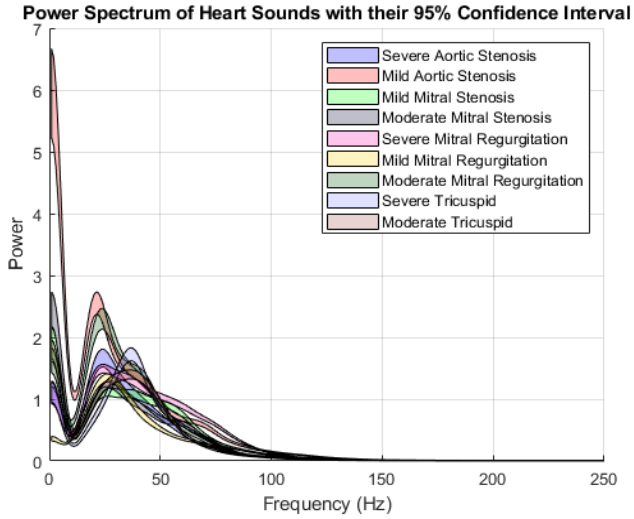

Figure 3. Power spectrum analysis of heart sounds with 95% confidence intervals

3.2. Training and validation performance

The model was trained and validated using a dataset of heart sound recordings across multiple diagnostic categories. This included different types of VHD as well as a healthy class for comparison. The model's performance was assessed through accuracy metrics, confusion matrices, and receiver operating characteristic (ROC) curves.

This confusion matrix as shown in Figure 4 illustrates the classification performance of the model during the training phase across ten disease categories, including severe aortic stenosis, mild mitral stenosis, and healthy subjects. The matrix shows the true positive rates and misclassification patterns for each target class. the model performed well in distinguishing severe cases of aortic stenosis with a high classification accuracy of 98.6% [22]. This confusion matrix summarizes the classification performance of the proposed CNN–BiLSTM model during the training phase across ten diagnostic classes. Each row represents the true class, while each column corresponds to the predicted class. High values along the main diagonal indicate strong class-wise recognition, with minimal confusion between unrelated valvular conditions. The few off-diagonal entries primarily occur between clinically similar disease severities, reflecting realistic diagnostic challenges rather than random errors.

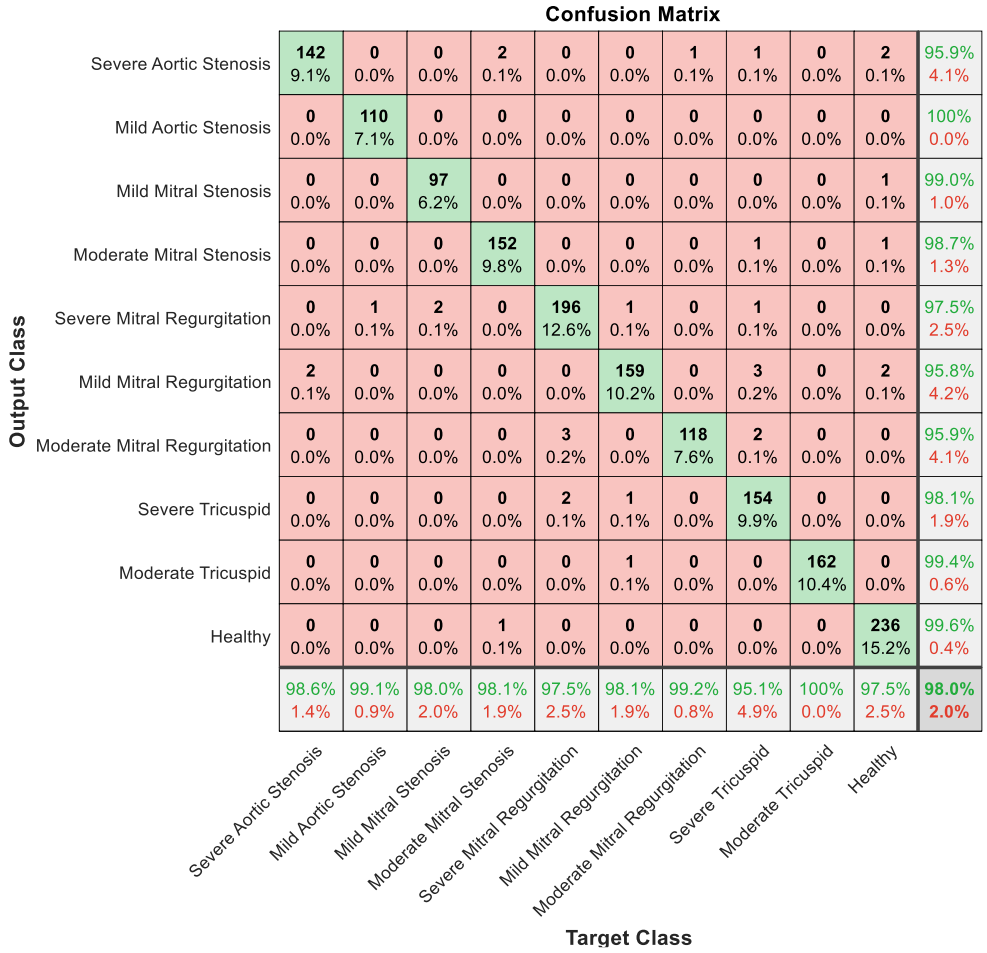

Figure 4. Confusion matrix for training folds of the VHD classification model

The ROC curves in Figure 5 demonstrate the diagnostic capability of the model during the training phase for each VHD category. This figure shows the ROC curves for each VHD category during the training phase. Each plot in Figure 5 corresponds to one diagnostic class and illustrates the trade-off between true positive rate and false positive rate. The consistently high area under the curve (AUC) values across all classes indicate excellent separability and strong learning capacity of the proposed model. These results confirm that the classifier achieves high sensitivity and specificity during training. These curves indicate the model's ability to separate the true positive rate from the false positive rate, with AUC values exceeding 0.95 for all categories.

Figure 6 presents the confusion matrix results from the testing phase, showcasing the model's performance in distinguishing between different VHD. This confusion matrix depicts the classification performance of the model on the independent testing dataset. The results demonstrate strong generalization, with high true positive rates maintained across all disease categories and healthy subjects. Misclassifications

are infrequent and mainly occur between adjacent severity levels of the same valvular pathology, such as mild versus moderate disease. The minimal confusion between healthy and pathological classes highlights the model's suitability for clinical screening applications. The model correctly classified 97.5% of severe aortic stenosis cases, with minimal false positives and negatives, indicating robust testing performance.

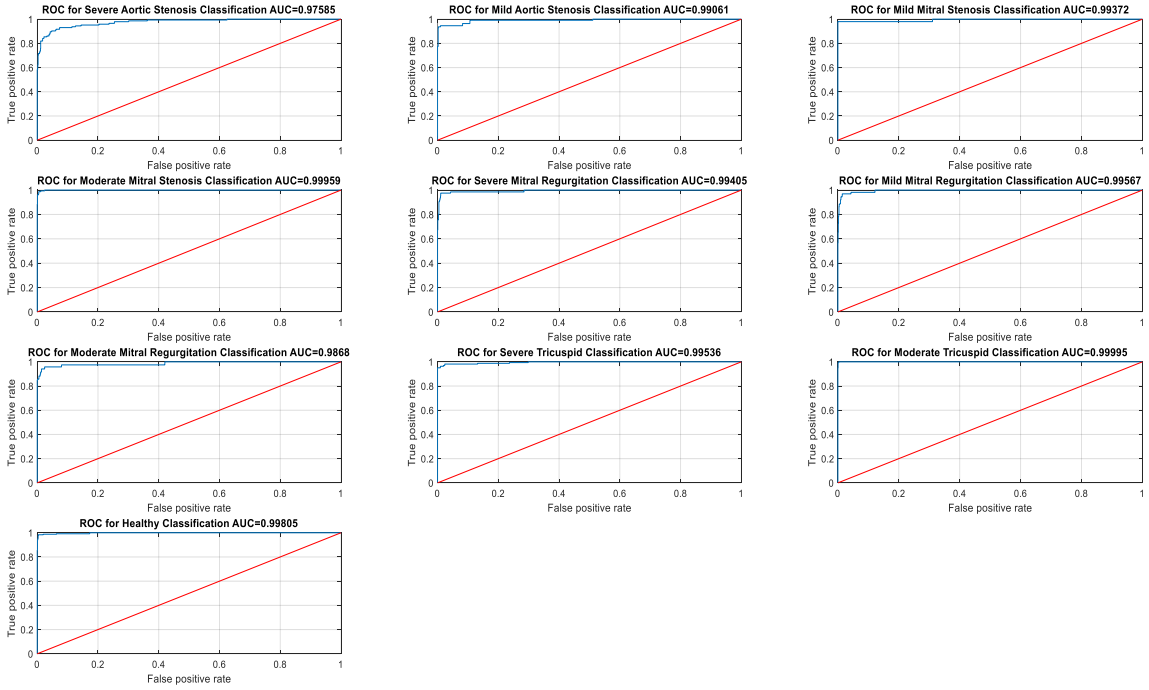

Figure 5. ROC curves for training folds of the VHD classification model

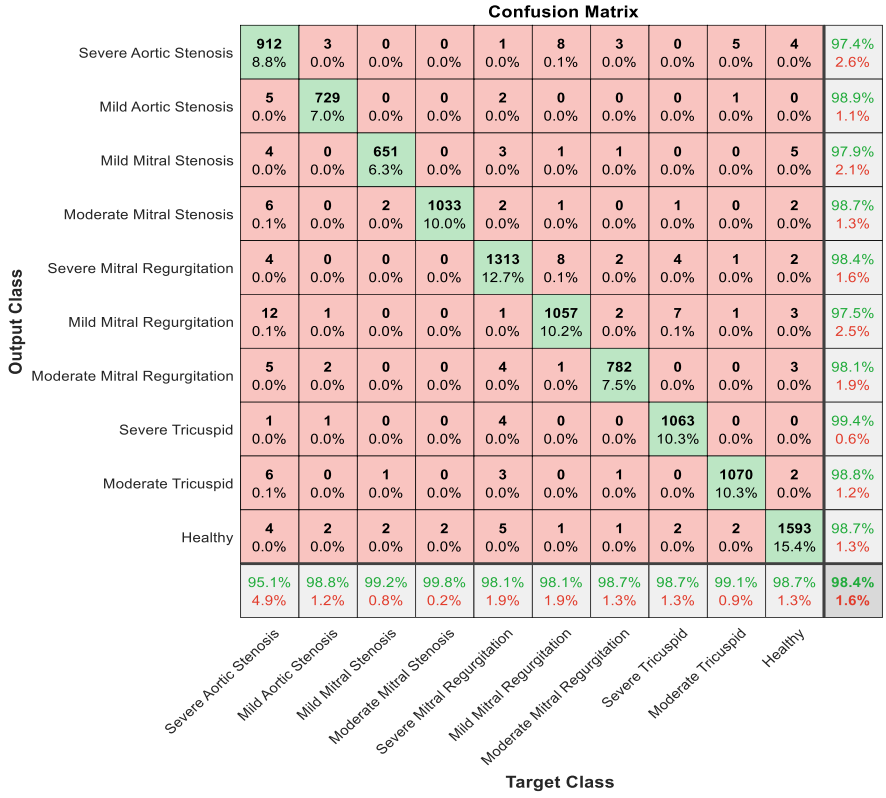

Figure 6. Confusion matrix for testing folds of the VHD classification model

The ROC curves in Figure 7 represent the model's diagnostic performance during testing. This figure presents the ROC curves for each diagnostic class during the testing phase, with each plot in Figure 7 corresponding to individual valvular conditions and the healthy class. The curves exhibit near-saturated behavior, with AUC values approaching 1.0 across all categories, indicating excellent diagnostic performance on unseen data. These results confirm that the proposed model retains high sensitivity and specificity beyond the training set. Importantly, the ROC analysis complements the confusion matrix by demonstrating robust discrimination even in a multi-class setting. High AUC values, approaching 1.0 across all categories, confirm that the model accurately distinguishes between healthy and diseased heart sound recordings.

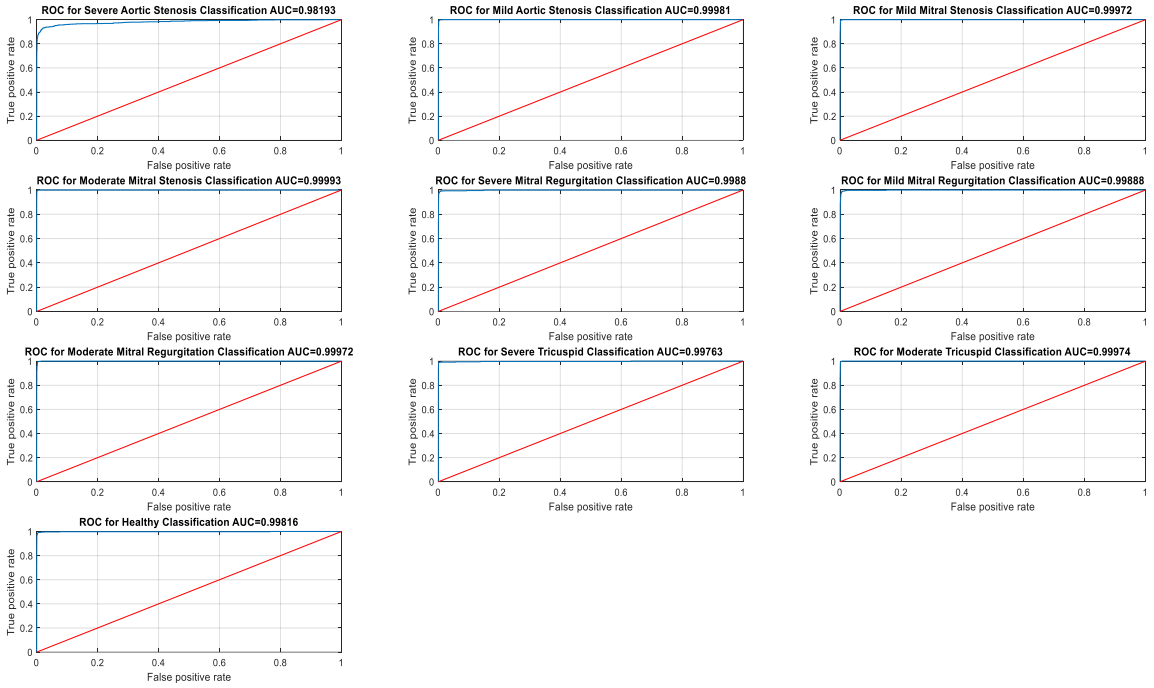

Figure 7. ROC Curves for testing folds of the VHD classification model

3.3. Discussion

This study presents a novel hybrid model that integrates CNN and BiLSTM networks for the classification of VHD using heart sound recordings. The proposed model achieved high performance, reflected by precision and recall metrics across different VHD categories. A 10-fold cross-validation was employed to evaluate the robustness of our model. Results showed consistently high classification accuracy across all disease categories, with the highest accuracy recorded for mild aortic stenosis (99.81%). The overall performance was strong, with accuracy levels surpassing 95% in all cases.

The confusion matrix revealed a low number of misclassifications, demonstrating the model's precision. For instance, severe aortic stenosis was classified correctly 98.6% of the time, and mild mitral stenosis achieved near-perfect classification at 99.9%. False positives and negatives were minimal, further reinforcing the model's reliability. ROC analysis provided additional support for the model's diagnostic capabilities, with all disease categories achieving high AUC values. The AUC for moderate mitral stenosis was 0.99993, making it one of the best-performing categories. Other conditions, like severe mitral regurgitation and severe tricuspid disease, also exhibited high AUC values, indicating strong discriminatory power between the different heart sound patterns. This performance ensures reliable classification in real-world scenarios. The results demonstrate superior performance compared to traditional machine learning models, which often rely on handcrafted features and do not fully exploit the potential of raw heart sound data. The integration of CNNs for feature extraction and BiLSTMs for temporal dependency analysis allows for a more comprehensive understanding of the heart sounds, enabling better classification performance across different types of VHD. The model's accuracy surpassed 98%, significantly improving upon prior work like Barua *et al.* [17], whose handcrafted feature-based model also showed strong performance but was limited by its feature extraction approach, which was not fully automated.

The success of this model points toward the feasibility of using AI-enhanced auscultation in clinical settings, especially in resource-limited environments where advanced diagnostic tools such as echocardiography may not be accessible. By utilizing electronic stethoscope data and applying minimal preprocessing, this method provides a cost-effective solution that could be deployed in portable devices, expanding access to early VHD detection globally. However, generalizability remains a challenge, as our model was primarily validated on inpatient data [23]. Further studies are needed to assess its performance in diverse clinical populations, including outpatient and remote settings [24], [25].

Despite the high accuracy, some limitations need to be addressed. The model exhibited reduced sensitivity for certain VHD subtypes, such as mixed valvular diseases, which may require further refinement of the architecture. Additionally, the dataset was relatively balanced across VHD classes, but real-world clinical data may present more skewed distributions, potentially affecting model performance. Future research should focus on improving sensitivity for complex VHD cases and validating the model across broader clinical populations. Moreover, incorporating additional data sources, such as patient history and other biomarkers, may enhance diagnostic accuracy and provide a more holistic approach to VHD diagnosis. To provide a more comprehensive evaluation beyond overall accuracy and ROC AUC, Table 3 reports class-wise precision, recall, and F1-score derived from the confusion matrix (Figure 6). These metrics offer deeper insight into the model's strengths and potential failure modes, particularly for clinically similar valvular conditions.

Table 3. Class-wise precision, recall, and F1-score for VHD classification

Class	Precision (%)	Recall (%)	F1-score (%)
Severe aortic stenosis	97.4	98.6	98
Mild aortic stenosis	98.8	99	98.9
Mild mitral stenosis	99.2	97.9	98.5
Moderate mitral stenosis	99.8	98.7	99.2
Severe mitral regurgitation	98.1	98.4	98.3
Mild mitral regurgitation	98.1	97.5	97.8
Moderate mitral regurgitation	98.7	98.1	98.4
Severe tricuspid disease	99.7	99.4	99.6
Moderate tricuspid disease	99	98.8	98.9
Healthy	98.9	98.7	98.8
Overall	98.8	98.4	98.5

Although the proposed CNN–BiLSTM model demonstrates consistently high precision and recall across all classes, the confusion matrix reveals a small number of structured misclassifications that provide insight into model limitations. The most frequent errors occur between adjacent severity levels of the same valve pathology, particularly mild versus moderate mitral and aortic valve diseases. These confusions are clinically plausible, as early-stage valvular abnormalities often exhibit overlapping acoustic signatures, making precise severity differentiation challenging even for expert clinicians. For example, mild mitral regurgitation is occasionally misclassified as moderate mitral regurgitation, suggesting that subtle spectral–temporal differences may not always be fully captured.

In contrast, severe valvular diseases, including severe aortic stenosis, severe mitral regurgitation, and severe tricuspid disease, exhibit very high recall (>98%), indicating that the model is highly sensitive to advanced pathological patterns. This is a desirable property in clinical screening applications, where missing severe disease carries a greater risk than mild misclassification. Misclassification between healthy subjects and pathological classes is minimal, with precision and recall for the healthy class exceeding 98%. This suggests strong discrimination between normal heart sounds and abnormal murmurs, reinforcing the model's suitability for early screening and triage.

ROC analysis as in Figure 6 further corroborates these findings, with all classes achieving near-saturated curves and AUC values close to 1.0. However, the presence of limited misclassifications despite extremely high AUC values highlights the importance of reporting class-wise precision, recall, and F1-score in multi-class clinical problems, as ROC curves alone may mask subtle yet clinically relevant errors. Overall, the failure cases identified are concentrated in clinically similar and acoustically overlapping disease categories, rather than across unrelated classes. This indicates that future improvements should focus on enhancing sensitivity to fine-grained severity differences, potentially through higher-resolution spectral features, attention mechanisms, or multimodal clinical data integration.

While the proposed ValveHealthNet framework demonstrated strong performance across all evaluated VHD categories, it should be noted that the current validation was primarily conducted on inpatient heart sounds recordings. As inpatient populations often exhibit clearer pathological signatures, model performance may differ in outpatient or community-based screening scenarios where murmurs can be subtler

and noise levels higher. Future work will therefore focus on prospective external validation using outpatient datasets and multi-center cohorts to assess robustness across diverse acquisition conditions and patient demographics. In particular, evaluating the model on heart sounds collected with different electronic stethoscopes and in ambulatory environments will be critical to ensure real-world applicability. Additionally, domain adaptation strategies such as fine-tuning with small labeled samples from new centers, noise-aware training, and data augmentation with realistic acoustic artifacts will be explored to enhance robustness against recording variability and environmental noise. These approaches are expected to improve generalization without substantially increasing model complexity.

4. CONCLUSION

This study introduced ValveHealthNet, a lightweight deep learning model specifically designed to enhance the detection and classification of VHD using heart sound recordings. The model demonstrated consistently high accuracy, precision, and recall across multiple VHD categories, highlighting its capability to differentiate between subtle acoustic signatures associated with different valve pathologies. A notable strength of ValveHealthNet is its minimal preprocessing requirements, which make it highly suitable for implementation on embedded systems and portable electronic stethoscopes. These characteristics position the model as a practical tool for resource-limited settings, potentially enabling early VHD screening in primary care clinics, rural hospitals, or community health programs where access to echocardiography is limited. Beyond its high classification performance, ValveHealthNet exhibits clinically meaningful behavior: it maintains excellent sensitivity for severe disease categories, while most misclassifications occur between adjacent disease severities, reflecting plausible diagnostic challenges that even human experts face. These results demonstrate that AI-assisted auscultation can complement traditional diagnostic workflows, providing a cost-effective and scalable solution for large-scale VHD screening, triage, and monitoring. Despite these strengths, several areas warrant further investigation. The model's current validation is primarily limited to inpatient heart sound recordings, which may differ from community or outpatient settings where murmurs are subtler and background noise levels are higher. Future research should therefore focus on external validation across diverse clinical environments and patient populations, including multi-center datasets and ambulatory recordings. Additionally, integrating complementary data sources such as patient history, echocardiography reports, and other biomarkers could enhance diagnostic precision and robustness, particularly for mixed or complex valvular diseases. Exploration of advanced techniques, such as attention mechanisms, multimodal fusion, and domain adaptation strategies, may further improve the model's sensitivity to fine-grained pathological differences without compromising computational efficiency. In summary, ValveHealthNet represents a promising step toward AI-enabled auscultation, offering a practical, accurate, and scalable approach for early detection and classification of VHD. With further validation and refinement, it has the potential to transform routine cardiac screening, improve patient outcomes through earlier diagnosis, and expand access to high-quality cardiovascular care globally.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ausilah Alfraihat	✓	✓	✓	✓	✓	✓			✓	✓				
Wafaa Al-Sharu	✓	✓								✓				
Ali Mohammad Alqudah	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The author declares that there are no known conflicts of interest associated with this publication. There are no financial or personal relationships that could inappropriately influence or bias the content of this work.

INFORMED CONSENT

Not applicable. This study did not involve human participants, human data, or any personally identifiable information. All data used were either publicly available, fully anonymized, or derived from non-human sources, and therefore no informed consent was required from individuals.

ETHICAL APPROVAL

Not applicable. This research did not involve human subjects, human biological materials, or experimental procedures on animals. The work was conducted solely on computational models, publicly available datasets, or non-sensitive data that did not require intervention with living organisms. Therefore, ethical approval from an institutional review board or animal ethics committee was not necessary for this study.

DATA AVAILABILITY

The data that support the findings of this study are available within the article at http://doi.org/10.1016/j.compbiomed.2022.105599, reference [17].

REFERENCES

[1] S. Queirós *et al.*, "Validation of a novel software tool for automatic aortic annular sizing in three-dimensional transesophageal echocardiographic images," *Journal of the American Society of Echocardiography*, vol. 31, no. 4, pp. 515-525, 2018, doi: 10.1016/j.echo.2018.01.007.

[2] S. Coffey, B. Cox, and M. J. A. Williams, "Lack of progress in valvular heart disease in the pre-transcatheter aortic valve replacement era: increasing deaths and minimal change in mortality rate over the past three decades," *American Heart Journal*, vol. 167, no. 4, pp. 562-567.e2, 2014, doi: 10.1016/j.ahj.2013.12.030.

[3] N. Frey *et al.*, "Symptoms, disease severity and treatment of adults with a new diagnosis of severe aortic stenosis," *Heart*, vol. 105, no. 22, pp. 1709–1716, 2019, doi: 10.1136/heartjnl-2019-314940.

[4] F. Li, Z. Zhang, L. Wang, and W. Liu, "Heart sound classification based on improved mel-frequency spectral coefficients and deep residual learning," *Frontiers in Physiology*, vol. 13, p. 1084420, 2022, doi: 10.3389/fphys.2022.1084420.

[5] P. A. McCullough *et al.*, "The healthcare burden of disease progression in medicare patients with functional mitral regurgitation," *Journal of Medical Economics*, vol. 22, no. 9, pp. 909–916, 2019, doi: 10.1080/13696998.2019.1621325.

[6] M. Thoenes *et al.*, "Patient screening for early detection of aortic stenosis (AS)-review of current practice and future perspectives," *Journal of Thoracic Disease*, vol. 10, no. 9, pp. 5584–5594, 2018, doi: 10.21037/jtd.2018.09.02.

[7] S. K. M. Gardezi *et al.*, "Cardiac auscultation poorly predicts the presence of valvular heart disease in asymptomatic primary care patients," *Heart*, vol. 104, no. 22, pp. 1832–1835, 2018, doi: 10.1136/heartjnl-2018-313082.

[8] J. Chen, W. Li, and M. Xiang, "Burden of valvular heart disease, 1990-2017: results from the global burden of disease study 2017," *Journal of Global Health*, vol. 10, no. 2, pp. 1–10, 2020, doi: 10.7189/jogh.10.020404.

[9] E. Donal, D. Muraru, and L. Badano, "Artificial intelligence and the promise of uplifting echocardiography," *Heart*, vol. 107, no. 7, pp. 523–524, 2021, doi: 10.1136/heartjnl-2020-318718.

[10] B. A. Virnig, N. D. Shippee, B. O'Donnell, J. Zeglin, and S. Parashuram, "Trends in the use of echocardiography, 2007 to 2011: Data Points #20," in *Data Points Publication Series*, Rockville (MD): Agency for Healthcare Research and Quality (US), 2011.

[11] S. Blecker *et al.*, "Temporal trends in the utilization of echocardiography in Ontario, 2001 to 2009," *JACC: Cardiovascular Imaging*, vol. 6, no. 4, pp. 515–522, 2013, doi: 10.1016/j.jcmg.2012.10.026.

[12] A. Papolos, J. Narula, C. Bavishi, F. A. Chaudhry, and P. P. Sengupta, "U.S. hospital use of echocardiography: insights from the nationwide inpatient sample," *Journal of the American College of Cardiology*, vol. 67, no. 5, pp. 502–511, 2016, doi: 10.1016/j.jacc.2015.10.090.

[13] W. A. Zoghbi *et al.*, "Recommendations for noninvasive evaluation of native valvular regurgitation: a report from the American society of echocardiography developed in collaboration with the society for cardiovascular magnetic resonance," *Journal of the American Society of Echocardiography*, vol. 30, no. 4, pp. 303–371, 2017, doi: 10.1016/j.echo.2017.01.007.

[14] E. Partovi, A. Babic, and A. Gharehbaghi, "A review on deep learning methods for heart sound signal analysis," *Frontiers in Artificial Intelligence*, vol. 7, 2024, doi: 10.3389/frai.2024.1434022.

[15] J. A. Lee and K. C. Kwak, "Heart sound classification using wavelet analysis approaches and ensemble of deep learning models," *Applied Sciences*, vol. 13, no. 21, 2023, doi: 10.3390/app132111942.

[16] L. O.-Reyes, M. A. A.-Arévalo, E. G.-Canseco, R. F. I.-Hernández, and R. C.-Galván, "A deep-learning approach to heart sound classification based on combined time-frequency representations," *Technologies*, vol. 13, 2025, doi: 10.3390/technologies13040147.

[17] P. D. Barua *et al.*, "An accurate valvular heart disorders detection model based on a new dual symmetric tree pattern using stethoscope sounds," *Computers in Biology and Medicine*, vol. 146, 2022, doi: 10.1016/j.compbiomed.2022.105599.

[18] Z. Jiang *et al.*, "Automated valvular heart disease detection using heart sound with a deep learning algorithm," *IJC Heart and Vasculature*, vol. 51, 2024, doi: 10.1016/j.ijcha.2024.101368.

[19] X. Li, G. A. Ng, and F. S. Schlindwein, "Transfer learning in heart sound classification using mel spectrogram," *2022 Computing in Cardiology (CinC)*, Tampere, Finland, 2022, pp. 1-4, doi: 10.22489/CinC.2022.046.

[20] E. M. Chambi, J. Cuela, M. Zegarra, E. Sulla, and J. Rendulich, "Benchmarking time-frequency representations of phonocardiogram signals for classification of valvular heart diseases using deep features and machine learning," *Electronics*, vol. 13, no. 15, 2024, doi: 10.3390/electronics13152912.
[21] J. Lee *et al.*, "Deep learning based heart murmur detection using frequency-time domain features of heartbeat sounds," *2022 Computing in Cardiology (CinC)*, Tampere, Finland, 2022, pp. 1-4, doi: 10.22489/CinC.2022.071.
[22] B. Walker, F. Krones, I. Kiskin, G. Parsons, T. Lyons, and A. Mahdi, "Dual Bayesian ResNet: a deep learning approach to heart murmur detection," in *Computing in Cardiology*, 2022, vol. 2022-September, pp. 1–4. doi: 10.22489/CinC.2022.355.
[23] R. Y. Mashhoor and A. Ayatollahi, "HeartSiam: a domain invariant model for heart sound classification," in *2022 8th International Iranian Conference on Signal Processing and Intelligent Systems, ICSPIS 2022*, 2022, pp. 1–5, doi: 10.1109/ICSPIS56952.2022.10044047.
[24] P. P. Sengupta, J. Kluin, S. P. Lee, J. K. Oh, and A. I. P. M. Smits, "The future of valvular heart disease assessment and therapy," *The Lancet*, vol. 403, no. 10436, pp. 1590–1602, 2024, doi: 10.1016/S0140-6736(23)02754-X.
[25] P. N. Waaler *et al.*, "Algorithm for predicting valvular heart disease from heart sounds in an unselected cohort," *Frontiers in Cardiovascular Medicine*, vol. 10, 2023, doi: 10.3389/fcvm.2023.1170804.

BIOGRAPHIES OF AUTHORS

Dr. Ausilah Alfraihat has a diverse academic background spanning several continents. As a biomedical engineer, she earned her Ph.D. from Drexel University, Philadelphia-United States, concentrating on thoracic vertebral morphology and the progression of scoliosis deformities. In her research, she employed advanced computational methodologies, utilizing machine learning models to understand the progression of deformities in scoliosis patients. Additionally, she completed her master's degree from the University of Malaya, where her research explored the potential of using saliva and urine samples to identify stages of breast carcinoma. She also holds a bachelor's degree from Jordan University of Science and Technology. She has shared her knowledge at various conferences and in several publications. Her international experiences have enriched her understanding and provided her with diverse perspectives in the realm of biomedical engineering. She can be contacted at email: ausilaha@hu.edu.jo.

Wafaa Al-Sharu received her M.Sc. from Jordan University of Science and Technology in 2008 and B.Sc. from Mutah University in 2001. Her research area is in the field of signal processing, image processing and analysis, deep learning, and machine learning. Currently she is an assistant lecturer at Department of Electrical Engineering, Hashemite University, Zarqa, Jordan. She can be contacted at email: wafaa.al-sharo3@hu.edu.jo.

Ali Mohammad Alqudah received a B.Sc. in Biomedical Systems Engineering and an M.Sc. in Computer Engineering from Yarmouk University, Jordan, in 2015 and 2018, respectively. He worked as a researcher and lab engineer at Yarmouk University between 2015 and 2022. Currently, he is a Ph.D. student and graduate research assistant in the Biomedical Engineering Program at the University of Manitoba, where he joined the Biomedical Instrumentation and Signal Analysis Lab. His research interests include biomedical signal processing and analysis, medical image processing and analysis, machine learning, deep learning, artificial intelligence in medicine, and obstructive sleep apnea. He can be contacted at email: ali_qudah@hotmail.com.